National Library
of Canada

Canadian Theses Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale
du Canada

Services des thèses canadiennes

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r.86/06)

Canada

THE UNIVERSITY OF ALBERTA

X-ray Crystallographic Studies on Subtilisins and Their Protein Inhibitors

by

©

Catherine Anne McPhalen

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF Doctor of Philosophy

Department of Biochemistry

EDMONTON, ALBERTA

FALL 1986

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-32487-2

Copyright Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Preliminary Crystallographic Data for the Serine Protease Inhibitor CI-2 from Barley Seeds" by C. A. McPhalen, C. Evans, K. Hayakawa, I. Jonassen, I. Svendsen, and M. N. G. James, published in *Journal of Molecular Biology* **168**: 445-447 (1983),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed ....John Kendrew........

Name ....Sir John Kendrew........

Editor-in-Chief, Journal of Molecular Biology

Date ....10th January 1986........

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Preliminary Crystallographic Data for the Serine Protease Inhibitor CI-2 from Barley Seeds" by C. A. McPhalen, C. Evans, K. Hayakawa, I. Jonassen, I. Svendsen, and M. N. G. James, published in *Journal of Molecular Biology* **168**: 445-447 (1983),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed ..... Claire Evans .....

Name ..... Claire Evans .....

Date ..... Feb. 6, 1986 .....

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Preliminary Crystallographic Data for the Serine Protease Inhibitor CI-2 from Barley Seeds" by C. A. McPhalen, C. Evans, K. Hayakawa, I. Jonassen, I. Svendsen, and M. N. G. James, published in *Journal of Molecular Biology* **168**: 445-447 (1983),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed K. Hayakawa

Name KOZO HAYAKAWA

Date Jan. 13, 1985

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Preliminary Crystallographic Data for the Serine Protease Inhibitor CI-2 from Barley Seeds" by C. A. McPhalen, C. Evans, K. Hayakawa, I. Jonassen, I. Svendsen, and M. N. G. James, published in *Journal of Molecular Biology* **168**: 445-447 (1983),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed Ib Jonassen

Name Ib JONASSEN

Date 6/1-1986

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Preliminary Crystallographic Data for the Serine Protease Inhibitor CI-2 from Barley Seeds" by C. A. McPhalen, C. Evans, K. Hayakawa, I. Jonassen, I. Svendsen, and M. N. G. James, published in *Journal of Molecular Biology* **168**: 445-447 (1983).

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed Ib. Svendsen

Name Ib Svendsen

Date January 6th, 1986

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Preliminary Crystallographic Data for the Serine Protease Inhibitor CI-2 from Barley Seeds" by C. A. McPhalen, C. Evans, K. Hayakawa, I. Jonassen, I. Svendsen, and M. N. G. James, published in *Journal of Molecular Biology* **168**: 445-447 (1983),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed ..M James..........

Name ..MICHAEL JAMES..........

Date ..Jan 14/86..........

University of Alberta
Edmonton
Canada T6G 2H7

Department of Biochemistry
474 Medical Sciences Building, Telephone (403) 432-2422

3 JAN. 1986

December 12, 1985

Elsevier Science Publishers B.V.
P.O. Box 211
1000 AE Amsterdam
The Netherlands

Dear Sir,

I am currently preparing my doctoral thesis in which I wish to include material previously published in FEBS Letters. To do so, I require a letter of copyright permission for the following manuscript:

C.A. McPhalen, H.P. Schnebli and M.N.G. James,
FEBS Letters (1985) 188, 55-58.

Parts of the manuscript may be included in a form similar to the published version, and the previous publication will be acknowledged.

Thank you for your help.

Sincerely,

C. McPhalen

Catherine McPhalen

CMcP/mw

PERMISSION GRANTED PROVIDED
ORIGINAL PUBLICATION IS MENTIONED
13 FEB 1986
FEDERATION [illegible]
BIOCHEMICAL SOCIETIES
COPYRIGHT HOLDER

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Crystal and molecular structure of the inhibitor eglin from leeches in complex with subtilisin Carlsberg"

by C. A. McPhalen, H. P. Schnebli and M. N. G. James,

published in *FEBS Letters* **188**: 55-58(1985),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed Hans P. Schnebli

Name Dr. H.P. Schnebli

Date Jan. 2. 1986

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Crystal and molecular structure of the inhibitor eglin from leeches in complex with subtilisin Carlsberg"

by C. A. McPhalen, H. P. Schnebli and M. N. G. James,

published in *FEBS Letters* **188**: 55-58(1985),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed MJames

Name MICHAEL JAMES

Date Jan 14 /86

Proceedings
OF THE
National Academy
of Sciences
OF THE UNITED STATES OF AMERICA

Officers of the Academy

Frank Press, *President*
James D. Ebert, *Vice President*
Bryce Crawford, Jr., *Home Secretary*
Walter A. Rosenblith, *Foreign Secretary*
Elkan R. Blout, *Treasurer*

Editorial Board of the Proceedings

Maxine F. Singer, *Chairman*

Robert A. Alberty
Paul Berg
Ronald Breslow
Edward E. David, Jr.
Stuart A. Kornfeld
Daniel E. Koshland, Jr.
Peter D. Lax
Robert E. Marshak
Daniel Nathans
Peter H. Raven
John Rodgers
Herbert E. Scarf
J. Edwin Seegmiller
Robert L. Sinsheimer
Solomon H. Snyder
Frank H. Westheimer

Managing Editor: Frances R. Zwanzig
Senior Associate Editor: Gary T. Cocks
Associate Editor: Cay Butler
Associate Editor: John M. Malloy
Associate Editor: Janet L. Morgan
Associate Editor: T. Pearson
Associate Editor: Dorothy P. Smith
Assistant Managing Editor: Joanne D'Amico

Senior Production Editor: Lynn A. Wassyng
Production Editors: Barbara A. Bacon, Ruth E. Crossgrove, Scott C. Herman, Peggy Leonard, Michael W. Neff, Janet L. Overton, Don C. Tippman, Deborah I. Weiner
Administrative Assistants: Delores Banks, Beaulah Edwards
Manuscript Coordinators: Cyndy Mathews, Jacqueline Perry
Circulation: Julia Little, Virginia Treadway

Editorial correspondence: Proceedings of the National Academy of Sciences, 2101 Constitution Avenue, Washington, DC 20418.

Business correspondence: Circulation Office of the Proceedings, National Academy of Sciences, 2101 Constitution Avenue, Washington, DC 20418.

Information for Contributors: See issue Number 1, January 1986.

Copyright: The National Academy of Sciences has copyrighted this journal as a collective work and does not own copyrights for individual articles. Requests for permission to reproduce parts of individual articles or for reprints of individual articles should be addressed to the authors. Microforms of complete volumes are available to regular subscribers only and may be obtained from University Microfilms, Xerox Corporation, Ann Arbor, MI 48103.

Subscriptions: All correspondence concerning subscriptions should be addressed to the Circulation Office of the Proceedings. Subscriptions are entered on a calendar year basis only. For 1986, subscription rates are as follows—in the United States: student/postdoctoral, $65; personal, $185; institutional, $215; elsewhere: student/postdoctoral, $100; personal, $220; institutional, $250. Subscribers are requested to notify the Circulation Office of the Proceedings 6 weeks in advance of any change of address; also the local postmaster. The Academy is not responsible for nonreceipt of issues because of an improper address unless a change of address is on file. The notice of address change should list both the old and new addresses. Claims for replacement copies will not be honored more than 60 days after the issue date for domestic subscribers and not more than 90 days after the issue date for foreign subscribers.

Back Issues: Volumes 78–82, January 1981 and thereafter, are available from the Circulation Office of the Proceedings. The price of a single issue is $21.00 for Volume 78 or $11.00 for Volumes 79–83.

Second class postage paid at Washington, DC, and at additional mailing offices.

PRINTED IN THE USA

Proceedings of the National Academy of Sciences of the United States of America (ISSN-0027-8424) *is published semimonthly by* The National Academy of Sciences, *2101 Constitution Avenue, Washington, DC 20418.*

© 1986 by The National Academy of Sciences of the United States of America.

POSTMASTER: Send address changes to: Proceedings of the National Academy of Sciences of the United States of America, 2101 Constitution Ave., Washington, DC 20418.

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin Novo"

by C. A. McPhalen, I. Svendsen, I. Jonassen, and M. N. G. James,

published in *Proceedings of the National Academy of Sciences* **82**: 7242-7246 (1985),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed ...Ib Svendsen...

Name ...Ib Svendsen...

Date ...January 6th, 1986...

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin Novo"

by C. A. McPhalen, I. Svendsen, I. Jonassen, and M. N. G. James,

published in *Proceedings of the National Academy of Sciences* **82**: 7242-7246 (1985),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed ....Ib Jonassen........

Name ....Ib JONASSEN........

Date ....6/1-1986........

Co-author Permission Form

Permission is hereby granted to Catherine A. McPhalen to use material from

"Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin Novo"

by C. A. McPhalen, I. Svendsen, I. Jonassen, and M. N. G. James,

published in *Proceedings of the National Academy of Sciences* **82**: 7242-7246 (1985),

in her thesis, entitled

X-ray Crystallographic Studies on Subtilisins and their Protein Inhibitors.

Signed ..... MJames .....

Name ..... MICHAEL JAMES .....

Date ..... Jan 14 /86 .....

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR	Catherine Anne McPhalen
TITLE OF THESIS	X-ray Crystallographic Studies on Subtilisins and Their Protein Inhibitors

DEGREE FOR WHICH THESIS WAS PRESENTED Doctor of Philosophy

YEAR THIS DEGREE GRANTED FALL 1986

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) C. McPhalen

PERMANENT ADDRESS:

Dept. of Biochemistry
University of Alberta
Edmonton, Alberta

DATED 24 April 1986

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled X-ray Crystallographic Studies on Subtilisins and Their Protein Inhibitors submitted by Catherine Anne McPhalen in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Michael James
Supervisor

....
Reinhart Reithmeier
....

Jane S. Richardson
External Examiner

Date 24 April 1986

To my parents, Barrie and Noreen McPhalen, and to my
husband, Bernard Lemire

Abstract

Three crystal structures of serine proteinase inhibitors, two complexed to serine proteinases, have been solved and refined at high resolution.

Subtilisin Carlsberg was the first bacterial serine proteinase to be discovered; it was isolated from *Bacillus subtilis*. Subtilisin Novo is a related bacterial serine proteinase, isolated from *Bacillus amyloliquefaciens*. Subtilisins Carlsberg and Novo have high sequence homology to each other, but no sequence homology to known eukaryotic serine proteinases. The subtilisins and the eukaryotic serine proteinases do share ~~a~~ similar geometry of the catalytic residues in the enzyme active site.

Chymotrypsin inhibitor 2 (CI-2) is a protein from the seeds of the Hiproly strain of barley. It inhibits a number of serine proteinases, including the subtilisins and α-chymotrypsin. CI-2 has been assigned to the potato inhibitor 1 (PI-1) family of serine proteinase inhibitors on the basis of sequence homologies. Eglin-c is another member of the PI-1 family, isolated from the leech *Hirudo medicinalis*. Eglin-c also inhibits a wide range of serine proteinases, including the subtilisins, α-chymptrypsin, mammalian leukocyte elastases, and cathepsin G. Neither CI-2 nor eglin-c contain disulphide bridges; such bridges are a common feature in other inhibitor families and are thought to stabilize and provide conformational rigidity for the reactive site loop.

The structure of a molecular complex of subtilisin Novo and CI-2 has been solved by the molecular replacement method and refined to a crystallographic *R*-factor of 0.154 for data in the range 8.0 to 2.1Å. CI-2 binds in the active site of subtilisin Novo in the manner of a good substrate, but is not cleaved. The folding of the polypeptide chain of the inhibitor is unlike the folding of inhibitors from other families with known structures.

The structure of a molecular complex between subtilisin Carlsberg and eglin-c has also been solved by the molecular replacement method and refined to an *R*-factor of 0.136 for data in the range 8.0 to 1.8Å. Overall, the structures of subtilisin Novo and subtilisin Carlsberg are very similar. Differences in the active site regions of the two enzymes indicate that there is a small but significant induced fit of enzyme to inhibitor upon complex formation. Differences in the structures of CI-2 and eglin-c are observed due to changes in the nature of their hydrophobic cores and to packing contacts in the crystal, but the overall folding of the inhibitors and the conformations of their reactive site loops are almost identical. Both inhibitors have an extensive network of hydrogen bonds and salt bridges from residues supporting and stabilizing the conformation of the reactive site loop to residues of the loop.

The structure of the inhibitor CI-2 in the absence of an enzyme has been solved by the molecular replacement method and refined to an *R*-factor of 0.198 for data in the range 8.0 to 2.0Å. The structure of the main body of the inhibitor is very similar to its structure in complex with subtilisin Novo, but the reactive site loop of the native inhibitor is a somewhat disordered region with high temperature factors. This kind of disorder in native inhibitors has been observed for other inhibitor families, and is believed to indicate flexibility and adaptability of the reactive site loop.

Evidence is presented that CI-2 and eglin-c act by the standard mechanism of Laskowski and Kato (*Ann. Rev. Biochem. 49*, 593-626, 1980) for protein inhibitors of serine proteinases. Comparison of CI-2 and eglin-c with members of other inhibitor families shows some conserved structural and electrostatic features that may contribute to the inhibitory nature of these small proteins.

Acknowledgements

During the time I have spent doing the work included in this thesis, a vast array of people have provided me with technical support, cheerful assistance in many matters, friendship, moral support, and intellectual stimulation. If I named them all here and gave them the full credit due to them, these acknowledgements would be longer than the rest of the thesis. I hope they all know how much I value all they have done for me.

I thank Mike James for making the time spent on my doctorate a period of tremendous learning and scientific excitement. He has shown me how to do science as carefully, thoroughly, clearly, and honestly as possible, and at the same time passed on some of his great enthusiasm and interest in discerning how nature functions.

The atmosphere in the lab during my time here has been exceptionally warm and close, and I regard the people I have worked with in the lab as a second family. Anita Sielecki has been a good friend and a thoughtful mentor on a personal level; she has also taught me a great deal about the careful and thorough refinement and analysis of protein structures. Randy Read and Masao Fujinaga have spent a large amount of time teaching me about the basics of protein crystallography and the inner workings of many of the local programs for processing and analyzing X-ray data. They have also done their best to make me computer-literate, and have shown tremendous patience in answering my endless questions. In

particular, Randy taught me much about heavy-atom techniques and protein refinement. Masao has constantly spurred me to think about scientific matters more closely and thoroughly, and has introduced me to the joys of long-distance backpacking. Osnat Herzberg has also taught me much about heavy-atom refinement techniques. John Moult has introduced me to the world of theoreticians and calculators, and shown me that even people without a strong mathematical bent can pursue those sorts of analyses without fear. Koto Hayakawa has been an absolutely essential part of the work in this thesis. She grew the crystals for all three structures described here, and cheerfully passed on some of her great store of wisdom on growing and handling crystals. Colin Broughton left the lab shortly after I arrived, but he left a legacy of invaluable computer programs and a computer graphics system that has been indispensible in my work. One final benefit of working in such a close and international lab is that I have been introduced to many new and exciting kinds of food from all parts of the world.

I have had many enjoyable associations with people from other labs in the department, both scientific and personal. In particular, Wayne Anderson has kept me aware of and interested in the world beyond serine proteinases. Dave Bacon has written computer programs that I use daily, and his efforts have also provided us with some lovely color graphics facilities.

Many thanks to my family, who have always encouraged me and supported me along my chosen path, even though the path has wandered at times. Most especially I thank my husband, Bernard Lemire, who has been a source of strength, intellectual stimulation, love, and friendship through these years.

This research has been supported by the Medical Research Council of Canada, through a grant to the MRC Group in Protein Structure and Function at the University of Alberta. I have been supported as a graduate student by Studentships from the Natural Sciences and Engineering Research Council of Canada, and the Alberta Heritage Foundation for Medical Research.

Table of Contents

Chapter Page

1. Introduction 1

  1.1 Serine Proteinases and Their Catalytic Mechanism 4

  1.2 The Subtilisins 13

  1.3 Protein Inhibitors of Serine Proteinases 19

  Bibliography 27

2. Subtilisin Novo and CI-2 33

  2.1 Structure Solution and Refinement 34

    2.1.1 Crystallization 34

    2.1.2 Data Collection and Processing 38

    2.1.3 Structure Determination 40

    2.1.4 Structure Fitting and Refinement 49

    2.1.5 Quality of the Refined Complex Structure ..59

  2.2 The Structure of Subtilisin Novo 75

    2.2.1 Secondary and Tertiary Structures 75

    2.2.2 Calcium Binding Sites 89

  2.3 The Structure of CI-2 93

  2.4 The Subtilisin Novo:CI-2 Complex 99

    2.4.1 Inhibitor Binding Interactions 99

    2.4.2 The Reactive Site Bond and the Active Site 106

    2.4.3 Solvent Structure of the Complex 111

    2.4.4 Accessibility of Protein Residues 119

  Bibliography 125

3. Subtilisin Carlsberg and Eglin-c 130

  3.1 Structure Solution and Refinement 131

    3.1.1 Crystallization 131

3.1.2 Data Collection and Processing 133
3.1.3 Structure Determination 134
3.1.4 Structure Fitting and Refinement 137
3.1.5 Quality of the Refined Complex Structure .143
3.2 The Structure of Subtilisin Carlsberg 156
3.2.1 Secondary and Tertiary Structures 156
3.2.2 Ion Binding Sites 164
3.3 The Structure of Eglin-c 168
3.4 The Subtilisin Carlsberg:Eglin-c Complex 173
3.4.1 Inhibitor Binding Interactions 173
3.4.2 The Reactive Site Bond and the Active Site ...177
3.4.3 Solvent Structure of the Complex 180
3.4.4 Accessibility of Protein Residues 186
Bibliography 190
4. Native CI-2 194
4.1 Structure Solution and Refinement 195
4.1.1 Crystallization 195
4.1.2 Data Collection and Processing 196
4.1.3 Structure Determination 198
4.1.4 Structure Fitting and Refinement 203
4.1.5 Quality of the Refined CI-2 Structure 209
4.2 The Structure of the Native CI-2 Inhibitor 221
4.2.1 Secondary and Tertiary Structures 221
4.2.2 Solvent Structure 225
4.2.3 Accessibility of Protein Residues 228
Bibliography 231
5. Discussion and Comparisons of Structures 233

5.1 Subtilisin Novo and Subtilisin BPN' 234
5.2 Subtilisin Carlsberg and Subtilisin Novo 236
    5.2.1 Global Comparison 236
    5.2.2 Active Site Regions 240
    5.2.3 Solvent Structure 243
5.3 Subtilisin Carlsberg and α-Chymotrypsin 246
5.4 CI-2 and Eglin-c 255
5.5 Inhibitor Interactions with Serine Proteinases .260
5.6 CI-2 Native and CI-2 in Complex 263
5.7 Why Are Eglin-c and CI-2 Good Inhibitors? 270
5.8 Future Directions 277
Bibliography 281

List of Tables

Table Page

2.1 Crystal Data for the Subtilisin Novo:CI-2 Complex 39

2.2 Course of Least-squares Refinement 54

2.3 Final Refinement Parameters and Results 60

2.4 Summary of Peptide Bond Geometry 68

2.5 Residues with Poor Electron Density at Cycle 71 73

2.6 Secondary Structural Elements of Subtilisin Novo 80

2.7 Hydrogen-bonding Parameters for Subtilisin Novo 81

2.8 Geometry of Calcium Ion Binding Sites 92

2.9 Secondary Structural Elements of CI-2 96

2.10 Hydrogen-bonding Parameters for CI-2 100

2.11 Intermolecular Contacts ≤4.0Å for Subtilisin Novo and CI-2 102

2.12 Hydrogen Bonds Between Subtilisin Novo and CI-2 105

2.13 Accessibilities by Residue Type 120

3.1 Crystal Data for the Subtilisin Carlsberg:Eglin-c Complex 134

3.2 Course of Least-squares Refinement 141

3.3 Final Refinement Parameters and Results 144

3.4 Summary of Peptide Bond Geometry 151

3.5 Residues with Poor Electron Density at Cycle 57 155

3.6 Secondary Structural Elements of Subtilisin Carlsberg 161

3.7 Hydrogen-bonding Parameters for Subtilisin Carlsberg 162

3.8 Geometry of Ion Binding Sites 167

Table Page

3.9 Secondary Structural Elements of Eglin-c 171

3.10 Intermolecular Contacts ≤4.0Å for Subtilisin Carlsberg and Eglin-c 174

3.11 Hydrogen Bonds Between Subtilisin Carlsberg and Eglin-c 176

3.12 Accessibilities by Residue Type 187

4.1 Crystal Data for Native CI-2 197

4.2 Course of Least-squares Refinement 205

4.3 Final Refinement Parameters and Results 210

4.4 Summary of Peptide Bond Geometry 216

4.5 Residues with Poor Electron Density at Cycle 133 220

4.6 Secondary Structural Elements of CI-2 223

4.7 Hydrogen-bonding Parameters for CI-2 226

5.1 Distances Between Equivalent Atoms, Subtilisin Carlsberg and α-Chymotrypsin 248

5.2 Hydrogen Bonds of the Catalytic Triad 251

5.3 Contacts (Å) in the Active Sites of Serine Proteinase Complexes 252

5.4 Hydrogen Bonds of Subtilisin:Inhibitor Complexes 262

List of Figures

Figure Page

1.1 Reaction Pathway for Serine Proteinase Catalysis 7

2.1 Amino Acid Sequences of Subtilisins Novo and Carlsberg 35

2.2 Amino Acid Sequences of CI-2 and Eglin-c 37

2.3 Rotation Function Map, Subtilisin Novo:CI-2 Complex 44

2.4 Translation Function Map, Subtilisin Novo:CI-2 Complex 47

2.5 Progress of the Subtilisin Novo:CI-2 Complex Refinement 53

2.6 σ_A Plot to Estimate Coordinate Error 63

2.7 Atomic Coordinate Error by the Method of Cruickshank 65

2.8 Variation in B-factor Along the Polypeptide Chain 66

2.9 ϕ-ψ Plot for Subtilisin Novo 69

2.10 ϕ-ψ Plot for CI-2 71

2.11 Regions of Good and Poor Electron Density 74

2.12 Views of Subtilisin Novo 77

2.13 Topology of Subtilisin Novo 78

2.14 Binding Sites for Ca^{2+} in Subtilisin Novo 90

2.15 Views of CI-2 95

2.16 Hydrogen Bonding of the Reactive Site Loop in CI-2 97

2.17 Bridging Water Molecules in the β-Sheet of CI-2 99

2.18 The Subtilisin Novo:CI-2 Complex 101

2.19 Interactions Between Subtilisin Novo and CI-2 103

2.20 Electron Density of the Scissile Bond in CI-2 108

Figure Page

2.21 Solvent Molecules Associated With the Complex 112

2.22 Residue Accessibilities in the Complex 114

2.23 Buried Water Channel in Subtilisin Novo 117

3.1 Progress of the Subtilisin Carlsberg:Eglin-c Complex Refinement 140

3.2 σ_A Plot to Estimate Coordinate Error 147

3.3 Atomic Coordinate Error by the Method of Cruickshank 149

3.4 Variation in *B*-factor Along the Polypeptide Chain 152

3.5 ϕ-ψ Plot for Subtilisin Carlsberg and Eglin-c 153

3.6 Regions of Good and Poor Electron Density 157

3.7 Views of Subtilisin Carlsberg 159

3.8 The *cis*-Peptide Bond Preceding Thr211 163

3.9 Ion Binding Sites in Subtilisin Carlsberg 166

3.10 Views of Eglin-c 169

3.11 Hydrogen Bonding of the Reactive Site Loop in Eglin-c 172

3.12 Interactions Between Subtilisin Carlsberg and Eglin-c 175

3.13 Electron Density of the Scissile Bond in Eglin-c 178

3.14 Solvent Molecules Associated With the Complex 180

3.15 Residue Accessibilities in the Complex 183

4.1 Progress of CI-2 Refinement 204

4.2 SDS Gel of CI-2 from Crystals 208

4.3 σ_A Plot to Estimate Coordinate Error 212

4.4 Atomic Coordinate Error by the Method of Cruickshank 213

Figure Page

4.5 Variation in *B*-factor Along the Polypeptide Chain 214

4.6 ϕ-ψ Plot for CI-2 217

4.7 Regions of Good and Poor Electron Density 219

4.8 Views of CI-2 222

4.9 Hydrogen Bonding of the Reactive Site Loop in Native CI-2 224

4.10 Solvent Molecules Associated With CI-2 227

4.11 Residue Accessibilities in CI-2 228

5.1 Superposition of Subtilisins Carlsberg and Novo 236

5.2 Superposition of Active Sites of Subtilisins 241

5.3 Superposition of Ca^{2+} Ion Binding Sites 245

5.4 Superposition of Subtilisin Carlsberg and α-Chymotrypsin Active Sites 249

5.5 Superposition of Eglin-c and CI-2 256

5.6 Superposition of Eglin-c and CI-2 Reactive Site Loops 257

5.7 Superposition of CI-2 Molecules 263

5.8 Difference *B*-Factor Plot, CI-2 Native vs CI-2 in Complex 265

5.9 Superposition of CI-2 Reactive Site Loops 266

5.10 Superposition of Enzyme:Inhibitor Complexes 269

5.11 Superposition of Eglin-c and OMTKY3 Reactive Site Loops 273

List of Symbols and Abbreviations

a, b, c	unit cell axes
B	thermal motion parameter
CI-2	chymotrypsin inhibitor 2 from barley seeds
DFP	diisopropylfluorophosphate
d_{min}	minimum interplanar spacing of diffraction data
e	electron
$\lvert\mathbf{E}\rvert$	normalized structure factor amplitude
$\lvert\mathbf{E}_c\rvert$	normalized calculated structure factor amplitude
$\lvert\mathbf{E}_o\rvert$	normalized observed structure factor amplitude
f	atomic scattering factor
$\mathbf{F}$	structure factor (vector quantity)
$\mathbf{F}_c$	calculated structure factor
$\mathbf{F}_o$	observed structure factor
$\lvert\mathbf{F}\rvert$	structure factor amplitude
$\lvert\mathbf{F}_o\rvert$	observed structure factor amplitude
$\lvert\mathbf{F}_c\rvert$	calculated structure factor amplitude
I	intensity
k_{cat}	rate constant for catalysis
K_M	Michaelis constant
m	figure of merit
m_c	m for calculated (model) phase
MIR	multiple isomorphous replacement
NMR	nuclear magnetic resonance
OMTKY3	domain 3 of turkey ovomucoid inhibitor

PI-1	potato inhibitor 1
PMSF	phenylmethanesulfonylfluoride
PSTI	pancreatic secretory trypsin inhibitor
PTI	pancreatic trypsin inhibitor
R	standard crystallographic residual
r	correlation coefficient
rms	root-mean-square
SDS	sodium dodecyl sulphate
SGPA	*Streptomyces griseus* Protease A
SGPB	*Streptomyces griseus* Protease B
SSI	*Streptomyces* subtilisin inhibitor
STI	soybean trypsin inhibitor
α_c	calculated structure factor phase
$\sigma(x)$	standard deviation in x
σ_A	phase probability parameter

---

1. Introduction

Proteins are fascinating and intricate devices designed by nature to perform a myriad of functions in the cell, ranging from provision of structural support through transport of metabolites to catalysis of biochemical reactions. The function and properties of a protein are determined essentially by its three-dimensional structure (the relative spatial arrangement of its amino acids and the folding of its polypeptide backbone). This structure can be visualized at near-atomic resolution by means of X-ray crystallography. Once the structure of a protein is known, the major challenge for the crystallographer is to relate that structure to other biochemical and physical data for the protein to determine how it carries out its function. This can be a long and difficult process; solving the crystal structure of a single protein can take many months or years, and to date the structures of only 200 to 250 different proteins have been solved. Hypotheses about structure-function relationships can be made on the basis of a few related structures, but are often difficult to test thoroughly (e.g. Baldwin, 1975).

Recent advances in genetic manipulation have extended the potential for understanding structure-function relationships in proteins. Site-directed mutagenesis may provide clues as to how small alterations in structure, such as single amino acid changes, are correlated with changes in functional properties (e.g. Alber *et al.*, 1986). X-ray crystallography is important in determining how the structure of a mutated protein has changed. The eventual goal of this kind of work is to design and synthesize proteins to perform new and desirable functions, often for use in industrial or pharmaceutical applications. Present knowledge of structure-function relationships is not sufficient to be able to design new proteins on a rational basis, although some work is being done on altering known proteins to improve function (Craik *et al.*, 1985).

Enzymes are proteins that act as catalysts for biological reactions that generally would not occur at perceptible rates in their absence. They accelerate reaction rates by factors of several orders of magnitude, and are specific in both the reaction that they catalyze and the substrates that they act upon. Precise knowledge of the structure of an enzyme is essential for understanding enzymatic catalysis. The special chemical environment that enables an enzyme to accelerate a chemical reaction is created by the spatial arrangement of the components of the enzyme. Crystallographic studies allow direct visualization of these spatial relations. At present, crystallography is unable to examine directly enzyme-substrate interactions during a reaction, because the time scale of an X-ray experiment is many orders of magnitude longer than that of an enzymatic reaction. Even if the X-ray data could be collected in milliseconds, arranging the molecules in the crystal to act in synchrony would be a major obstacle to obtaining an ordered view of a reaction. In spite of this limitation, changes in the structure of an enzyme upon interaction with inhibitors and analogs of intermediates in the reaction pathway may be seen crystallographically, and can provide important information about the mechanism of catalysis.

Crystallographic studies of two enzyme:inhibitor complexes and one native inhibitor structure form the body of this thesis. Chapter 2 discusses the solution of the structure of the complex between subtilisin Novo and the protein proteinase inhibitor CI-2 from barley seeds. The structures of the enzyme and the inhibitor are described in detail, as well as the interactions between them. Chapter 3 is a similar discussion of the structure of the complex between subtilisin Carlsberg and eglin-c, an inhibitor from leeches. Chapter 4 describes the structure of the native inhibitor CI-2, crystallized in the absence of an enzyme. Chapter 5 contains comparisons of the structures of the two enzymes, the two inhibitors from the complexes, the interactions in the two complexes, and the native versus the complexed inhibitors. Some comparisons are also made with other serine proteinases and other families of protein inhibitors of serine proteinases.

For the non-crystallographer, an excellent introduction to X-ray crystallography can be found in chapter 6 of "Proteins: Structures and Molecular Properties" by T. E. Creighton (1983). Two other standard texts in the area are

"Protein Crystallography" by T. L. Blundell and L. N. Johnson (1976), and "Crystals, X-rays and Proteins" by D. Sherwood (1976). In a discussion such as this, is it difficult to give a brief but meaningful description of the subject to someone not already familiar with it.

1.1 Serine Proteinases and Their Catalytic Mechanism

Proteinases are a group of enzymes that function biologically to catalyze the hydrolysis of peptide bonds and thus cleave other proteins or peptides. The serine proteinases are distinguished as a class by having an unusually reactive serine residue in the enzyme active site that plays an important role in catalysis. The active site also contains a histidine and an aspartic acid residue that are essential to catalysis; these three residues together are termed the catalytic triad. Two general families of serine proteinases are known at present; the chymotrypsin-like enzymes and the subtilisins. Members of both these families use the same mechanism of catalysis, but the two families differ completely in three-dimensional structure except in the arrangement of their catalytic residues. Chymotrypsin-like serine proteinases have been found in a wide variety of organisms, both prokaryotic and eukaryotic, but to date the subtilisins have been identified only in prokaryotes. Because their overall three-dimensional structures are so different, these two families of serine proteinases are believed to have developed through convergent evolution.

A wide variety of biological functions are performed by serine proteinases (Neurath, 1984). They hydrolyze proteins in food in the digestive systems of many organisms. They can activate other proteins by selective proteolytic cleavage e.g. most steps in the blood clotting cascade require that the serine proteinase generated in one step activates the serine proteinase zymogen in the next. Many serine proteinases are extra-cellular and play a degradative and protective role for the cells that secrete them. Numerous other biological roles also have been attributed to these enzymes, ranging from involvement in virus replication (Jacobson & Baltimore, 1968) to differentiation of adipocytes (Cook *et al.*, 1985).

Serine proteinases have been studied extensively, both crystallographically and by many other physical and biochemical techniques. The kinetic data on serine proteinases have been reviewed by Bender & Killheffer (1973) for the chymotrypsin-like enzymes, and by Philipp & Bender (1983) for the subtilisins. The X-ray crystallographic structures of a number of serine proteinases have been solved, including subtilisin Novo (Wright *et al.*, 1969; Drenth *et al.*, 1972), and many of the structures of the chymotrypsin-like enzymes have been refined at high resolution. Reviews by Huber & Bode (1978) and Kraut (1977) contain discussions of these structural results and their relation to the chemical and kinetic data. Aspects of the catalytic mechanism studied by both crystallographic and NMR techniques have been reviewed by Steitz & Shulman (1983). James *et al*. (1980) present a detailed proposal for the catalytic mechanism of serine proteinases that is based on the proposal of Kraut (1977) and further structural studies on enzyme:product and :inhibitor complexes.

The transition state theory of reaction rates is generally used to describe how enzymes are able to act as catalysts. This theory proposes that the rate of a chemical reaction will be accelerated if the transition state of the reaction can be reached more easily. Enzymes are thought to bind more strongly to the transition state than to the substrate for the reaction, increasing the concentration of the transition state and thus accelerating the reaction (Pauling, 1946; Jencks, 1966; Wolfenden, 1972; Lienhard, 1973). Serine proteinases catalyze acyl transfer reactions. The transition state for this type of reaction is believed to resemble a tetrahedral adduct to the acyl group; such a tetrahedral adduct is an intermediate in the reaction pathway for acylation and deacylation (Jencks, 1969). Thus serine proteinases should be designed to stabilize a transition state closely resembling this tetrahedral intermediate.

The pathway of chemical reactions believed to be followed by serine proteinases during catalysis is shown in Figure 1.1. The following discussion of the molecular events in catalysis is based heavily on the mechanistic proposal of James *et al*. (1980). The substrate is assumed to

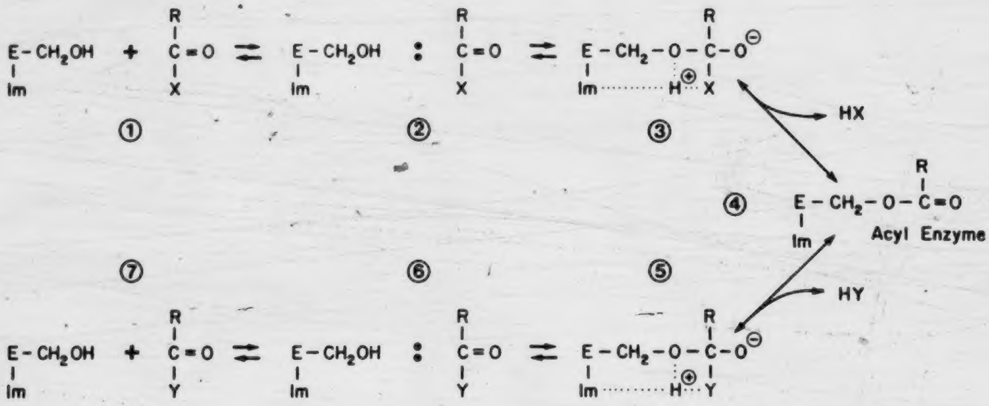


Figure 1.1. Reaction Pathway for Serine Proteinase Catalysis (after James *et al*. (1980)). The upper line shows the sequence of reactions in the formation of the acyl enzyme; the lower line shows deacylation. The enzyme is indicated as E, with the active site serine and histidine residues as $-CH_2OH$ and Im respectively. RCOX is the substrate, which can be a peptide, ester or amide. Non-covalent Michaelis complexes of the enzyme and substrate are depicted in (2) and (6), covalent tetrahedral intermediates in (3) and (5), and the acyl enzyme in (4).

be a protein or peptide. The initial step in the reaction is the formation of a non-covalent complex between the enzyme and substrate, termed the Michaelis complex. Models of such a complex are difficult to find for crystallographic purposes but, as may be seen in later chapters, the protein inhibitors of the serine proteinases may provide the best model yet available. Tight, productively-oriented binding of the substrate is promoted by complementarity of regions on the surfaces of enzyme and substrate close to the enzyme active site. The second step of the reaction is the formation of a covalently bound tetrahedral intermediate. The carbonyl oxygen atom of the P_1 amino acid residue in the substrate[1] lies in a strongly polarizing environment, in a pocket on the enzyme surface known as the oxyanion hole (Henderson, 1970; Robertus *et al*., 1972b). The polarization withdraws electrons from the carbonyl carbon atom of P_1 and induces on it a partial positive charge. The proximity of this charge to the O^{γ} of the active site serine of the enzyme facilitates transfer of the proton from the serine to the leaving group, and the formation of a bond between O^{γ} and the substrate carbonyl carbon. The path followed by the proton to the leaving group is determined by the special electronic environment existing in the active site region. This special environment is created by the presence of the serine O^{γ}, the histidine sidechain (hydrogen bonded to the essential aspartic acid residue of the enzyme), the oxyanion hole, and the carbonyl carbon - carbonyl oxygen dipole of the substrate. Neutron diffraction studies have shown that a deuteron is present on the histidine $N^{\epsilon 2}$ in a model of the tetrahedral intermediate (Kossiakoff & Spencer, 1981); this is presumed to be the proton being passed from the serine O^{γ} to the P_1' leaving group. As the P_1 carbonyl carbon forms a bond with the serine O^{γ}, it must rehybridize from sp^2 to sp^3

---

[1] The notation of Schechter and Berger (1967) is used to facilitate discussion of the interactions between a proteinase and bound peptides. Amino acid residues of substrates are numbered P_1, P_2, etc., in the amino-terminal direction and P_1', P_2', etc. in the carboxy-terminal direction from the scissile bond. The complementary subsites of the enzyme binding region are numbered S_1, S_2 and S_1', S_2', etc.

to form the tetrahedral intermediate. This lengthens and thus weakens the peptide bond.

The third step in the reaction sequence is the resolution of the tetrahedral intermediate into a covalent acyl enzyme complex. The leaving group, the amide nitrogen of the P_1' residue, accepts the proton from the histidine $N^{\epsilon 2}$ of the enzyme, breaking the peptide bond and releasing the C-terminal part of the substrate from the enzyme. This leaves the N-terminal part of the substrate bound to the enzyme via an ester linkage, which was originally presumed to be planar. Spectroscopic data (Bernhard & Lau, 1971) are not consistent with the existence of a planar ester, which would be very stable chemically and difficult to hydrolyze at the catalytic rates observed for these enzymes. Instead, James *et al.* (1980) have proposed that the carbonyl carbon of the substrate retains its pyramidal conformation and its partial positive charge in the acyl enzyme. This leaves the acyl enzyme poised for the deacylation steps in the reaction, which is probably the reverse of the acylation by the principle of microscopic reversibility.

In the deacylation reaction, a water molecule takes the place of the nitrogen atom of the departed C-terminal portion of the substrate. The partial positive charge on the acyl enzyme induces the formation of a strongly nucleophilic hydroxyl ion from the water molecule; the histidine side-chain may play a role in orienting the incoming water molecule (Kossiakoff & Spencer, 1981). The hydroxyl ion attacks the carbonyl carbon of the substrate. A second tetrahedral intermediate is formed from the acyl enzyme, with both the attacking hydroxyl group and the serine O^{γ} bound to the carbonyl carbon, and the proton in transit between the two nucleophiles. This tetrahedral intermediate breaks down to a non-covalent enzyme:product complex. The proton is transferred to the serine O^{γ} as the bond to the substrate carbonyl carbon is broken, and the hydroxyl ion is incorporated into the newly formed carboxy-terminus of the cleaved substrate. The amino-terminal part of the substrate is then free to dissociate from the enzyme, leaving the enzyme active site fully regenerated and ready to begin another cycle.

Extensive chemical and kinetic evidence supports the existence of various intermediates in the reaction pathway for serine proteinases given here. A summary of this evidence is given by Kraut (1977) in his review of the catalytic mechanism of serine proteinases.

The mechanistic proposal outlined above addresses some of the issues that have been discussed for many years regarding the nature and role of each of the residues comprising the catalytic triad. Some of the questions that have been asked are: Why is the serine such a strong nucleophile? What are the roles of the histidine and the aspartic acid? The 'charge-relay' system of Birktoft & Blow (1972) was one attempt to explain the nucleophilicity of the serine. Early unrefined X-ray structures of α-chymotrypsin

(Blow *et al.*, 1969) and subtilisin (Wright *et al.*, 1969) were interpreted as showing a good hydrogen bond between the serine O^{γ} and the histidine $N^{\epsilon 2}$. The role of the negatively charged aspartic acid was to withdraw the proton from the serine O^{γ}, with the histidine as an intermediary, and thus render the serine more nucleophilic. In subsequently solved structures of serine proteinases, the serine-histidine hydrogen bond was not found; the O^{γ} to $N^{\epsilon 2}$ distances found range from 2.9Å in the kallikrein structure (Bode *et al.*, 1983) to 3.7Å in the refined structure of subtilisin (Matthews *et al.*, 1977). The angle between donor and acceptor in these structures is about 110°, which would give a severely distorted hydrogen bond. This distorted geometry is found in structures done at both high and low pH (Matthews *et al.*, 1977; Blevins & Tulinsky, 1985; Fujinaga *et al.*, 1985; James *et al.*, 1980; Bode & Schwager, 1975), so the distortion is not a function of the protonation state of the histidine.

According to James *et al.* (1980), the serine is not intrinsically more nucleophilic than usual, but its nucleophilicity is induced by the presence of the partial positive charge on the carbonyl carbon of the scissile bond. This partial positive charge arises from enhanced polarization of the bond between the carbonyl carbon and the carbonyl oxygen, due to proximity to the oxyanion hole. As well, structures of serine proteinases with analog[illegible] Michaelis complex or the tetrahedral intermediate show an excellent hydrogen bond between serine and histidine (e.g. Huber & Bode, 1978; Read *et al.*, 1983; Kossiakoff & Spencer, 1981), in contrast to the free enzyme. The formation of this bond upon binding of a substrate may enhance the nucleophilicity of the serine in the Michaelis complex.

The coupled histidine and aspartic acid residues have been assigned the role of 'proton shuttle' during catalysis by Kraut (1977), with the histidine accepting a proton from the serine and passing it to the leaving group during acylation, and reversing the sequence during deacylation. Although the pK_a of the histidine is around 7 in the free enzyme (Robillard & Shulman, 1974; Bachovchin & Roberts, 1978), Kossiakoff & Spencer (1981) have postulated that the burial of the histidine side-chain upon substrate binding may raise its pK_a to above 9.5 to allow it to extract protons from attacking nucleophiles. The specific role of the aspartic acid is then to orient the histidine side-chain for its interactions during catalysis, to maintain His in a single tautomeric form when $N^{\epsilon 2}$ is unprotonated in the free enzyme, and to provide ion pair stabilization when His is protonated in the enzyme-substrate complex. Moult *et al.* (1985) have performed quantum mechanical calculations that show that the electrostatic environment provided by the aspartic acid polarizes the bond between the $N^{\epsilon 2}$ of the histidine and the proton accepted from the serine, and may thus also enhance proton transfer.

1.2 The Subtilisins

Subtilisins are a group of extracellular bacterial serine proteinases that are most active at alkaline pH. They are excreted by a variety of strains of *Bacillus* as part of a developmental sequence that results in the bacteria producing heat-resistant spores. Studies of a deletion mutant of *Bacillus subtilis* I168 have shown that the proteinase is probably not necessary for spore development, but it may play a nutrient scavenging role in the conditions of nutrient deprivation necessary to initiate spore formation (Stahl & Ferrari, 1984). As a separate family of serine proteinases, the subtilisins differ greatly from the chymotrypsin-like enzymes in primary sequence and overall tertiary structure, have much broader substrate specificities, and have different pH optima. Members of the subtilisin family are highly homologous in sequence, and have very similar physical and enzymatic properties.

The X-ray structure of a fungal enzyme, proteinase K, has been solved recently (Pähler *et al.*, 1984); the folding of its main-chain is very similar to that of the subtilisins. Comparisons between proteinase K and the subtilisins will be of interest when the amino acid sequence of the fungal enzyme is known and its structure has been refined. This is the only member of the subtilisin family from a species other than *Bacillus* to have its three-dimensional structure solved, and few other non-*Bacillus* subtilisins are known.

Subtilisin Carlsberg was discovered by Linderstrøm-Lang and Ottesen in 1947, and was subsequently isolated and purified in the Carlsberg laboratories in Denmark (Linderstrøm-Lang & Ottesen, 1947; Güntelberg & Ottesen, 1954). The biological source of this enzyme was reported to be a strain of *Bacillus subtilis*. The purification of subtilisin BPN' was reported by Hagihara *et al*. (1958) and that of subtilisin Novo by Ottesen & Spector (1960). These two subtilisins were proved by subsequent work to be chemically identical (Olaitan *et al*., 1968; Robertus *et al*., 1971; Drenth *et al*., 1972). Originally, they both were reported to have been isolated from *B. subtilis*, but Welker and Campbell (1967) later showed that the organism producing BPN' was actually *B. amyloliquefaciens*. Other subtilisins have been purified from *B. amylosaccharíticus* (Tsuru *et al*., 1966), *B. licheniformis*, and *B. pumilis* (Keay & Moser, 1969). Keay & Moser (1969) tested the catalytic and immunological properties of subtilisins from several types of *Bacillus* and divided them into two groups: those that resembled subtilisin Novo and those that resembled subtilisin Carlsberg. The Novo group includes BPN', NRRLB3411, and *amylosaccharíticus*. The Carlsberg group includes *licheniformis* and *pumilis*. The Novo type has a lower ratio of esterase to proteinase activity, and the two types are not immunologically cross-reactive. Much of the early work on the isolation and characterization of subtilisins has been reviewed by Ottesen & Svendsen (1970). The chemical and physical properties of these enzymes were reviewed by Markland & Smith (1971), and the kinetics of subtilisin Novo by Philipp & Bender (1983).

The genes for the subtilisins from *B. amyloliquefaciens* and *B. licheniformis* have been cloned and expressed in *B. subtilis* (Wells *et al.*, 1983; Jacobs *et al.*, 1985). Both genes have been sequenced and the amino acid sequences of the proteins deduced from the DNA sequences. These deduced sequences show some differences from the published protein sequences. In the case of subtilisin Novo (the *B. amyloliquefaciens* sequence), the electron density map for the enzyme structure described in this thesis is consistent with the DNA sequence. The *B. licheniformis* sequence is that of a Carlsberg-type enzyme; the bacterial strain carrying the gene was isolated with antibodies raised against commercial subtilisin Carlsberg. The electron density map for the Carlsberg structure described here is consistent with the original protein sequence in all but one position. It is entirely possible that the protein used in these X-ray studies was produced by a *Bacillus* variant other than *B. licheniformis*, thus the differences between the DNA and protein sequences may be real.

The DNA sequences of the two cloned subtilisins both have a large open reading frame between the presumed ribosome binding site and the beginning of the mature protein sequence. The amino acid sequence deduced for the first 30 to 35 residues of this reading frame are homologous to signal sequences observed for other secreted proteins from Gram positive bacteria. Following the proposed signal peptide, there are approximately 75 amino acids before the start of the mature protein. Eukaryotic secreted proteinases are often synthesized in an inactive zymogen form with a 'pro-peptide' interfering with productive substrate binding, to prevent unwanted proteolysis inside the cell. This pro-peptide is removed once the proteinase has reached its extra-cellular destination. This phenomenon had been observed only for streptococcal proteinase in prokaryotes (Liu & Elliot, 1971). The proposed pro-peptide sequences for these subtilisins do have some sequence homology with known eukaryotic pro-peptides (Wells *et al.*, 1983).

The mature subtilisin Novo is composed of a single polypeptide chain of 275 amino acids, M_r = 27583 (Olaitan *et al.*, 1968). Subtilisin Carlsberg has 274 amino acid residues, M_r = 27292 (Smith *et al.*, 1968). The amino acid sequences of these two enzymes are given in Figure 2.1 (chapter 2), aligned on the basis of three-dimensional structural homology. Neither enzyme contains cysteine residues. Both subtilisins are very acid labile, but are highly resistant to denaturing agents such as 6*M* urea and detergents (Markland & Smith, 1971). This latter property has encouraged great interest in subtilisins by manufacturers of enzyme-based laundry detergents. Subtilisin Carlsberg is produced in greater quantities than any other enzyme, primarily for industrial use (Aunstrup *et al.*, 1979). A great deal of work is currently in progress on modifying subtilisins to improve heat stability and susceptibility to oxidation, using cloning and site-specific mutagenesis technology. Subtilisin BPN' (≡ subtilisin Novo) has been shown to be stabilized by Ca^{2+} and other salts (Matsubara *et al.*, 1958); this is characteristic of both prokaryotic and eukaryotic serine proteinases.

The X-ray crystallographic structures of subtilisin Novo (Drenth *et al.*, 1972) and subtilisin BPN' (Wright *et al.*, 1969) have been solved in two different crystal forms. This work provided part of the evidence that the two enzymes are chemically identical. The Novo structure was solved at 2.8Å resolution, with the enzyme active site acylated with diisopropylfluorophosphate (DFP). The BPN' structure was originally solved at 2.5Å resolution, with the enzyme active site acylated with phenylmethanesulfonylfluoride (PMSF). The structure of the native enzyme was later solved from difference maps, at a resolution of 2.0Å (Alden *et al.*, 1971), and refined by unrestrained difference Fourier methods to a crystallographic *R*-factor[a] of 0.229 (Matthews *et al.*, 1975). Atomic coordinates for the 2.8Å Novo structure and the 2.5Å BPN' structure have been deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977), but coordinates for the refined 2.0Å BPN' structure have not been published. The BPN' structure at various stages of refinement has been used to examine the enzyme structure with virtual substrates (Robertus *et al.*, 1972a; Robertus *et al.*, 1972b) and transition state analogs (Matthews *et al.*, 1975; Poulos *et al.*, 1976) bound in the active site, by means of difference Fourier maps (coefficients $|F_O|-|F_C|, \alpha_C$).

---

[a] The *R*-factor is a measure of agreement between observed and calculated structure factors. It is defined as $R=\Sigma||\mathbf{F}_o|-|\mathbf{F}_c||/\Sigma|\mathbf{F}_o|$.

The work described in this thesis was undertaken to study interactions between subtilisins and their protein inhibitors. The structures of subtilisin Novo and subtilisin Carlsberg from these enzyme:inhibitor complexes have been extensively refined by least-squares methods, and atomic coordinates for these refined structures will be deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). Since the original subtilisin Novo and BPN' structures were solved, advances in computer technology and improvements in refinement strategies and programs have allowed more thorough refinement of protein structures. A well-refined enzyme structure is essential in studying an enzymatic reaction mechanism in any detail; the precise placement of catalytically active groups, and the careful definition of possible hydrogen-bonding or electrostatic interactions are important in interpreting such a mechanism. More confidence may be placed in the atomic positions of a refined structure; the errors in those positions typically are around 1.0Å in an unrefined structure, whereas the errors in the refined atomic positions of structures from well-refined structures are around 0.15Å (James & Sielecki, 1983; Read *et al.*, 1983; Fujinaga *et al.*, 1985).

The structure of subtilisin Carlsberg is of particular interest because it had not been solved previously. Subtilisin Novo and subtilisin Carlsberg are 70% identical in sequence (Fig. 2.1) and sequence changes between the two are generally conservative or lie on the surface of the known subtilisin Novo molecule (Wright *et al.*, 1969). On this basis, and because of the similarities in the physical and biochemical properties of the two enzymes, their three-dimensional structures are also expected to be very similar. Even so, quantitative differences have been observed in some kinetic parameters and in the substrate specificities of Novo and Carlsberg (Barel & Glazer, 1968).

1.3 Protein Inhibitors of Serine Proteinases

Small proteins that act as inhibitors of serine proteinases have been isolated from many biological sources, from bacteria to plants and humans, and from many types of tissues or fluids. Their obvious biological function is to prevent undesirable proteolysis, but their physiological significance in many situations is not clear. One likely role is the prevention of premature activation of zymogens of digestive proteinases in the vertebrate pancreas. Another is the limitation of proteolytic cascades in processes such as blood clotting and complement activation. In plants, serine proteinase inhibitors are thought to protect against wounding and infestation by insects. Most inhibitors are discovered on the basis of their ability to inhibit standard screening enzymes; in the words of Laskowski & Kato (1980)

> They thus become proteins in search of a function, rather than proteins isolated to account for a previously discovered biological function. ...The lack of knowledge of true target enzymes for the majority of inhibitors is now one of the major stumbling blocks in the understanding of inhibitor evolution and inhibitor specificity.

Laskowski & Kato (1980) have divided the known serine proteinase inhibitors into more than 10 families, based on sequence homologies and patterns of disulphide bridge topology. Members of six of these families have been studied by X-ray crystallography; structures of some of the inhibitors are known both in the native state and complexed to an enzyme (Read & James, 1986). Many of the known inhibitors act by a standard mechanism (Laskowski & Kato, 1980). Kinetic and biochemical data, combined with the crystallographic results, have shown that these inhibitors bind to an enzyme in the manner of a good substrate, but are cleaved at a very slow rate. The inhibition results primarily from tight binding of the inhibitor to the enzyme (low K_M), slow hydrolysis (low k_{cat}), and slow release of the cleaved inhibitor.

The simplest form of the reaction between enzyme and inhibitor is

$$E + I \rightleftharpoons E \cdot I \rightleftharpoons E + I*$$

where E is the enzyme, I is the uncleaved inhibitor, E·I is the enzyme:inhibitor Michaelis complex and I* is the cleaved inhibitor (Finkenstadt & Laskowski, 1967). Three equilibrium constants can be defined for this reaction:

$$K_{assoc} = [E \cdot I]/([E][I + I^*])$$

$$K_{hyd} = [I^*]/[I]$$

$$K_a = [E \cdot I]/([E][I])$$

K_{assoc} should be high for a good inhibitor, in order that a large amount of the enzyme:inhibitor complex be present at equilibrium. K_{hyd} is the equilibrium constant for the hydrolysis of inhibitor, and K_a is the equilibrium constant for the association of enzyme and uncleaved inhibitor. K_{assoc} can be defined in terms of the other two constants as

$$K_{assoc} = K_a/(1 + K_{hyd})$$

(Read & James, 1986, Read *et al*., 1983). To maximize K_{assoc}, a good inhibitor should have a high K_a and a low K_{hyd}.

A high value of K_a implies strong favorable interactions between enzyme and inhibitor to promote formation of E·I. Crystal structures of inhibitors and enzyme:inhibitor complexes have shown that free inhibitors maintain conformations of their reactive site regions complementary to the enzyme active site (the reactive site is the region containing the bond cleaved when I is converted to I*). This relatively rigid conformational complementarity allows the inhibitor to bind to the enzyme with little loss of motional freedom, and thus little entropy cost relative to a conformationally labile substrate (Huber *et al.*, 1974; Sweet *et al.*, 1974; Blow, 1974). Consequently, the inhibitor has a higher binding energy than a substrate; K_a values for enzyme:inhibitor complexes are typically on the order of 10^{10} M^{-1}.

Finkenstadt *et al.* (1974) have demonstrated that K_{hyd} for most inhibitors is close to unity at neutral pH, whereas substrate hydrolysis rapidly goes to completion. Rigidity of the inhibitor reactive site region may contribute to this slow rate of hydrolysis by not permitting relaxation of the inhibitor after cleavage. Features such as disulphide bridges or hydrogen bonding networks encompassing the reactive site bond are present in all inhibitors of known structure; these features would hold together the two halves of a cleaved inhibitor and promote the reversal of the cleavage reaction. The net effect of the reactive site rigidity would be to decrease [I*] and increase [I], thus decreasing K_{hyd} overall.

In addition to high K_a and low K_{hyd}, a third factor has been suggested as a limit to inhibitor hydrolysis. Early studies showed that the rate of complex formation from enzyme and cleaved inhibitor was much slower than that from enzyme and intact inhibitor. These data were interpreted as indicating that a barrier to the hydrolysis of the inhibitor existed i.e. that E + I could form E·I much more rapidly than E + I* (Quast *et al.*, 1978). Some interactions in

inhibitor structures have been noted that could hinder the formation of the tetrahedral intermediate, a necessary step in hydrolysis after E + I has formed E·I (Fujinaga *et al.*, 1982; Read *et al.*, 1983). Recent data, however, show that the relative rates of complex formation from cleaved and intact inhibitor vary over several orders of magnitude for the reaction of one inhibitor with different enzymes (Ardelt & Laskowski, 1985). The observed interactions that could hinder the formation of the tetrahedral intermediate may be one factor in the slow hydrolysis of inhibitors, but there must be other factors that are dependent on differences in the enzyme structures or in their interactions with inhibitors. The known serine proteinase structures are very similar in their active site regions, so the differences may be quite subtle. Read & James (1986) have suggested two possibilities:
1) that the dynamic properties of the enzymes result in some complexes being more rigid and less reactive than others, and 2) that the S_1' region of the enzymes differs more than other parts of the active site and this could alter binding interactions with the P_1' residue, particularly after cleavage.

Inhibitors of serine proteinases have been the subject of the majority of crystallographic studies of inhibitors that are proteins as well. These inhibitors possess characteristics of good substrates such as the appropriate amino acid sequence around the reactive site for specific recognition by the enzyme, and they bind productively to the enzyme as a good substrate would, yet they are inhibitors, not substrates. The goal of the crystallographic work has been the identification of structural features that contribute to the inhibitory properties discussed in the preceding paragraphs. The crystal structures of protein inhibitors of serine proteinases that have been solved to date are reviewed in Read & James (1986).

The crystal structures of two new serine proteinase inhibitors are discussed in this thesis. CI-2, isolated from barley seeds, and eglin-c, from leeches, are both members of the potato inhibitor 1 family in the classification of Laskowski & Kato (1980). The archetype of this family is the potato chymotrypsin inhibitor 1 (PI-1), isolated from *Solanum tuberosum* by Ryan & Balls (1962). PI-1 is a strong inhibitor of α-chymotrypsin and subtilisin. It contains only one intrachain disulphide bridge that can be reduced and alkylated without affecting its inhibitory properties (Plunkett & Ryan, 1980). Eglin-c and CI-2 were assigned to this family of inhibitors on the basis of sequence homologies; neither contains disulphide bridges or cysteine residues.

CI-2 was originally isolated as the protein contributing most to the high lysine content of the albumin fraction of the Hiproly strain of barley (Jonassen, 1980). The amino acid sequence of this protein was determined; it consists of 83 amino acids, with a blocked amino-terminus (M_r = 9250) (Svendsen *et al*., 1980a) (Fig. 2.2, chapter 2). It was shown by immunochemical techniques to be identical to an inhibitor of α-chymotrypsin and microbial serine proteinases also isolated from barley seeds, and assigned to the PI-1 family of inhibitors (Svendsen *et al*., 1980b). The inhibitor reactive site bond was demonstrated to lie between Met59I and Glu60I[3] for interaction with α-chymotrypsin, subtilisin Novo and subtilisin Carlsberg (Jonassen & Svendsen, 1982).

Eglin-c was isolated from the leech *Hirudo medicinalis* on the basis of its anti-chymotryptic activity (Seemüller *et al*., 1977). It was subsequently found to form strong complexes ($K_i \cong 10^{-11}$ M) with granulocytic elastase and cathepsin G (Seemüller *et al*., 1980; Schnebli *et al*., 1985). It is very stable to denaturation by acid and heat and to proteolytic degradation, and does not react with any mammalian proteinases tested other than the ones mentioned above and α-chymotrypsin. All these properties have made it a potential candidate for therapeutic treatment of conditions such as emphysema and septicemia. Eglin-c consists of 70 amino acids, M_r = 8092, and its reactive site bond is located between Leu47I and Asp48I (Leu59I and Asp60I in the CI-2 numbering used here, Fig. 2.2, chapter 2). The gene for eglin-c has been synthesized and expressed in *E. coli* by recombinant DNA techniques by Rink *et al*. (1984). The recombinant eglin-c differs from the natural product only in being acetylated at the N-terminus; it retains all the biological properties of the natural product.

---

[3] An 'I' follows the sequence numbers of inhibitor residues to distinguish them from those of the enzymes.

When work was begun on the X-ray structures of CI-2 and eglin-c, the serine proteinase inhibitors of known structure all contained at least one disulphide bridge encompassing the reactive site bond. This disulphide is believed to contribute to the stability and rigidity of the reactive site loop of these kinds of inhibitors, and thus to their inhibitory properties (Laskowski & Kato, 1980; Read *et al*., 1983). The structures of CI-2 and eglin-c consequently were of particular interest to determine how they were stabilized enough without this covalent link to function by the standard inhibitory mechanism (if they did function by the standard mechanism). Since eglin-c has potential therapeutic applications, knowledge of its structure may eventually be useful in the design of drugs to inhibit specific kinds of harmful proteolysis. In addition, the X-ray structures of two enzymes, subtilisin Novo and subtilisin Carlsberg, and these two inhibitors, all of which lack disulphide bridges and all of which are particularly stable to denaturation, will be of interest to those concerned with the general structural basis of protein stability.

Bibliography

Alber, T., Daopin, S., Wilson, K., Bell, J., Wozniak, J., Cook, S., & Matthews, B. W. (1986) *Biophysical Journal 49*, 439a.

Alden, R. A., Birktoft, J. J., Kraut, J., Robertus, J. D., & Wright, C. S. (1971) *Biochem. Biophys. Res. Commun. 45*, 337-344.

Ardelt, W. & Laskowski, M. Jr. (1985) *Biochemistry 24*, 5313-5320.

Aunstrup, K., Anderson, O., Falck, E. A., & Nielsen, T. K. (1979) in *Microbial Technology* Vol. 1, eds. H. J. Peppler & D. Terlman, New York: Academic Press, 281-309.

Bachovchin, W. W. & Roberts, J. D. (1978) *J. Amer. Chem. Soc. 100*, 8041-8047.

Baldwin, J. M. (1975) *Prog. Biophys. Mol. Biol. 29*, 225-320.

Barel, A. O. & Glazer, A. N. (1968) *J. Biol. Chem. 243*, 1344-1348.

Bender, M. L. & Killheffer, J. V. (1973) *CRC Crit. Rev. Biochem. 1*, 149-199.

Bernhard, S. A. & Lau, S.-J. (1971) *Cold Spring Harbor Symp. Quant. Biol. 36*, 75-83.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. J. Jr., Brice, M. D., Rogers, J. K., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535-542.

Birktoft, J. J. & Blow, D. M. (1972) *J. Mol. Biol. 68*, 187-240.

Blevins, R. A. & Tulinsky, A. (1985) *J. Biol. Chem. 260*, 4264-4275.

Blow, D. M., Birktoft, J. J., & Hartley, B. S. (1969) *Nature (London) 221*, 337-340.

Blow, D. M. (1974) in *Bayer Symp. V, Proteinase Inhibitors*, eds. H. Fritz, H. Tschesche, L. J. Greene, & E. Truscheit, Berlin: Springer-Verlag, 677-678.

Blundell, T. L. & Johnson, L. N. (1976) *Protein Crystallography*, New York: Academic Press.

Bode. W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., & Bartunik, H. (1983) *J. Mol. Biol. 164*, 237-282.

Bode, W. & Schwager, P. (1975) *J. Mol. Biol. 98*, 693-717.

Cook, K. S., Groves, D. L., Min, H. Y., & Spiegelman, B. M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6480-6484.

Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R. J., & Rutter, W. J. (1985) *Science 228*, 291-297.

Creighton, T. E. (1983) *Proteins: Structures and Molecular Principles*, New York: W. H. Freeman & Co.

Drenth, J., Hol, W. G. J., Jansonius, J. N., & Koekoek, R. (1972) *Eur. J. Biochem. 26*, 177-181.

Finkenstadt, W. R. & Laskowski, M. Jr. (1967) *J. Biol. Chem. 242*, 771-773.

Finkenstadt, W. R., Hamid, M. A., Mattis, J. A., Schrode, J., Sealock, R. W., Wang, D., & Laskowski, M. Jr. (1974) in *Bayer Symp. V, Proteinase Inhibitors*, eds. H. Fritz, H. Tschesche, L. J. Greene, & E. Truscheit, Berlin: Springer-Verlag, 389-411.

Fujinaga, M., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1985) *J. Mol. Biol. 183*, 479-502.

Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M. Jr., & James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868-4872.

Güntelberg, A. V. & Ottesen, M. (1954) *Compt. Rend. Trav. Lab. Carlsberg, Ser. Chim. 29*, 36-48.

Hagihara, B., Matsubara, H., Nakai, M., & Okunuki, K. (1958) *J. Biochem. (Tokyo) 45*, 185-194.

Henderson, R. (1970) *J. Mol. Biol. 54*, 351-354.

Huber, R. & Bode, W. (1978) *Acc. Chem. Res. 11*, 114-122.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., & Steigemann, W. (1974) *J. Mol. Biol. 89*, 73-101.

Jacobs, M., Eliasson, M., Uhlén, M., & Flock, J.-I. (1985) *Nucleic Acids Research 13*, 8913-8926.

Jacobson, M. F. & Baltimore, D. (1968) *Proc. Natl. Acad. Sci. U.S.A. 61*, 77-84.

James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol. 163*, 299-361.

James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J., & Bauer, C.-A. (1980) *J. Mol. Biol. 144*, 43-88.

Jencks, W. P. (1966) in *Current Aspects of Biochemical Energetics*, eds. N. O. Kaplan & E. P. Kennedy, New York: Academic Press, 273-298.

Jencks, W. P. (1969) *Catalysis in Chemistry and Enzymology*, New York: McGraw-Hill, 463-552.

Jonassen, I. (1980) *Carlsberg Res. Commun. 45*, 47-58.

Jonassen, I. & Svendsen, I. (1982) *Carlsberg Res. Commun. 47*, 199-203.

Keay, L. & Moser, P. W. (1969) *Biochem. Biophys. Res. Commun. 34*, 600-604.

Kossiakoff, A. A. & Spencer, S. A. (1981) *Biochemistry 20*, 6462-6474.

Kraut, J. (1977) *Ann. Rev. Biochem. 46*, 331-358.

Laskowski, M. Jr. & Kato, I. (1980) *Ann. Rev. Biochem. 49*, 593-626.

Lienhard, G. E. (1973) *Science 180*, 149-154.

Linderstrøm-Lang, K. & Ottesen, M. (1947) *Nature 159*, 807-808.

Liu, T. Y. & Elliot, S. D. (1971) in *The Enzymes, 3rd ed.* Vol. 3, ed. P. Boyer, New York: Academic Press, 609-647.

Markland, F. S. & Smith, E. (1971) in *The Enzymes, 3rd ed.* Vol. 3, ed. P. Boyer, New York: Academic Press, 561-608.

Matsubara, H., Hagihara, B., Nakai, M., Komaki, T., Yonetani, T., & Okunuki, K. (1958) *J. Biochem. (Tokyo) 45*, 251-255.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., & Kraut, J. (1975) *J. Biol. Chem. 250*, 7120-7126.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., & Kraut, J. (1977) *J. Biol. Chem. 252*, 8875-8883.

Moult, J., Sussman, F., & James, M. N. G. (1985) *J. Mol. Biol. 182*, 555-566.

Neurath, H. (1984) *Science 224*, 350-357.

Olaitan, S. A., DeLange, R. J., & Smith, E. L. (1968) *J. Biol. Chem. 243*, 5296-5301.

Ottesen, M. & Spector, A. (1960) *Compt. Rend. Trav. Lab. Carlsberg 32*, 63-74.

Ottesen, M. & Svendsen, I. (1970) in *Meth. Enzymol.* Vol. XIX, eds. L. Lorand & G. Perlman, New York: Academic Press, 199-215.

Pähler, A., Banerjee, A., Dattagupta, J. K., Fujiwara, T., Lindner, K., Pal, G. P., Suck, D., Weber, G., & Saenger, W. (1984) *EMBO Journal 3*, 1311-1314.

Pauling, L. (1946) *Chem. Eng. News 24*, 1375-1377.

Philipp, M. & Bender, M. L. (1983) *Mol. & Cell. Biochem. 51*, 5-32.

Plunkett, G. & Ryan, C. A. (1980) *J. Biol. Chem. 255*, 2752-2755.

Poulos, T. L, Alden, R. A., Freer, S. T., Birktoft, J. J., & Kraut, J. (1976) *J. Biol. Chem. 251*, 1097-1103.

Quast, U., Engel, J., Steffen, E., Tschesche, H., & Kupfer, S. (1978) *Eur. J. Biochem. 86*, 353-360.

Read, R. J. & James, M. N. G. (1986) in *Proteinase Inhibitors*, ed. A .J .Barrett, Amsterdam: Elsevier, in press.

Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420-4433.

Rink, H., Liersch, M., Sieber, P., & Meyer, F. (1984) *Nucleic Acids Research 12*, 6369-6387.

Robertus, J. D., Alden, R. A., & Kraut, J. (1971) *Biochem. Biophys. Res. Commun. 42*, 334-339.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., & Wilcox, P. E. (1972a) *Biochemistry 11*, 2439-2449.

Robertus, J. D., Kraut, J., Alden, R. A., & Birktoft, J. J. (1972b) *Biochemistry 11*, 4293-4303.

Robillard, G. & Shulman, R. G. (1974) *J. Mol. Biol. 86*, 519-540.

Ryan, C. A. & Balls, A. K. (1962) *Proc. Natl. Acad. Sci. U.S.A. 48*, 1839-1844.

Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157-162.

Schnebli, H. P., Seemüller, U., Fritz, H., Maschler, R., Liersch, M., Bodmer, J. L., Virca, G. D., Lucey, E. C., Stone, P. G., & Snider, G. L. (1985) in *Intracellular Protein Catabolism*, eds. E. A. Khairallah, J. S. Bond, & J. W. C. Bird, New York: Alan Liss, Inc., 287-290.

Seemüller, U., Eulitz, M., Fritz, H., & Strobl, A. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1841-1846.

Seemüller, U., Meier, M., Ohlsson, K., Müller, H.-P., & Fritz, H. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1105-1117.

Sherwood, D. (1976) *Crystals, X-rays and Proteins*, London: Longman Group Ltd.

Smith, E. L., DeLange, R. J., Evans, H., Landon, M., & Markland, F. S. (1968) *J. Biol. Chem. 243*, 2184-2191.

Stahl, M. L. & Ferrari, E. (1984) *J. Bacteriol. 158*, 411-418.

Steitz, T. A. & Shulman, R. G. (1982) *Ann. Rev. Biophys. Bioeng. 11*, 419-444.

Svendsen, I., Jonassen, I., Hejgaard, J., & Boisen, S. (1980b) *Carlsberg Res. Commun. 45*, 389-395.

Svendsen, I., Martin, B., & Jonassen, I. (1980a) *Carlsberg Res. Commun. 45*, 79-85.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., & Blow, D. M. (1974) *Biochemistry 13*, 4212-4228.

Tsuru, D., Kira, H., Yamamoto, T., & Fukumoto, J. (1966) *Agr. Biol. Chem. (Tokyo) 30*, 1261-1268.

Welker, N. E. & Campbell, L. L. (1967) *J. Bacteriol. 94*, 1124-1130.

Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., & Chen, E. Y. (1983) *Nucleic Acids Research 11*, 7911-7925.

Wolfenden, R. (1972) *Acc. Chem. Res. 5*, 10-18.

Wright, C. S., Alden, R. A., & Kraut, J. (1969) *Nature (London) 221*, 235-242.

2. Subtilisin Novo and CI-2[1]

Subtilisin Novo is a bacterial serine proteinase active at alkaline pH. The three-dimensional X-ray crystallographic structure of this enzyme was solved by Drenth *et al.* (1972). The structure of a chemically identical enzyme, subtilisin BPN' was solved by Wright *et al.* (1969), and has been refined to a crystallographic *R*-factor[2] of 0.229 at 2.0Å resolution with no constraints on bond lengths and angles (Matthews *et al.*, 1977). The coordinates for the unrefined structures of both these enzymes are available in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). Subtilisin Novo has no sequence homology with the other family of serine proteinases, those resembling α-chymotrypsin. Its three-dimensional structure is also entirely different from the chymotrypsin-like enzymes except for the side-chains of the catalytic residues in its active site region.

CI-2 is a member of the potato inhibitor 1 family of serine proteinase inhibitors. It is isolated from the seeds of the Hiproly strain of barley (Svendsen *et al.*, 1980b). Members of this family of inhibitors show sequence homology with other inhibitor families only in the reactive site region. CI-2 is the first member of the PI-1 family to have had its X-ray structure solved.

---

[1] A version of portions of this chapter has been published [McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242-7246].

[2] The *R*-factor is a measure of agreement between observed and calculated structure factors. It is defined as $R=\Sigma||\mathbf{F}_o|-|\mathbf{F}_c||/\Sigma|\mathbf{F}_o|$.

Crystals of a molecular complex of subtilisin Novo and CI-2 have been grown. The molecular replacement method (Rossmann, 1973) was used to obtain an initial electron density map for the complex. It was possible to trace in this map all of the polypeptide backbone of the enzyme and part of the backbone of the inhibitor. The remainder of the CI-2 molecule (beginning at Leu20I) was fitted to the electron density after several cycles of restrained least-squares refinement of the complex structure. The refinement of the structure was completed after 71 cycles, at an *R*-factor of 0.154 for data in the resolution range from 8.0 to 2.1Å. The secondary and tertiary structures of the enzyme and the inhibitor have been examined in detail, as well as the interactions between the two in the complex.

2.1 Structure Solution and Refinement

2.1.1 Crystallization

Purified lyophilized CI-2 was prepared at the Carlsberg Research Center, Copenhagen, Denmark (Jonassen, 1980; Svendsen *et al.*, 1980a). During purification procedures, the inhibitor is subject to hydrolysis at 3 bonds near its amino-terminus, and the product with a 'ragged' amino-terminus was used in crystallization trials. Subtilisin Novo was a gift from Novo Industry, Bagsvaerd, Denmark

										10										20					
Carlsberg	A	Q	T	V	P	Y	G	I	P	L	I	K	A	D	K	V	Q	A	Q	G	F	K	G	A	N
Novo	A	Q	S	V	P	Y	G	V	S	Q	I	K	A	P	A	L	H	S	Q	G	Y	T	G	S	N
					T	T	T	T	T	T	T	T	α	α	α	α	α	α	T	T			T	T	T
						α	α	α	α	α	α					T	T	T							

					30										40										50
Carlsberg	V	K	V	A	V	L	D	T	G	I	Q	A	S	H	P	D	L	N	V	V	G	G	A	S	F
Novo	V	K	V	A	V	I	D	S	G	I	D	S	S	H	P	D	L	K	V	A	G	G	A	S	M
	T	β	β	β	β	β	β				T	T	T	T	T	T	T				β	β	β	β	β
	β																								

										60										70					
Carlsberg	V	A	G	E		A	Y	N	T	D	G	N	G	H	G	T	H	V	A	G	T	V	A	A	L
Novo	V	P	S	E	T	N	P	F	Q	D	N	N	S	H	G	T	H	V	A	G	T	V	A	A	L
	T	T	T	T									α	α	α	α	α	α	α	α	α	α	α	α	

					80										90										100
Carlsberg	D	N	T	T	G	V	L	G	V	A	P	S	V	S	L	Y	A	V	K	V	L	N	S	S	G
Novo	N	N	S	I	G	V	L	G	V	A	P	S	A	S	L	Y	A	V	K	V	L	G	A	D	G
										T	T	T	T	β	β	β	β	β	β	β		T	T	T	T

										110										120					
Carlsberg	S	G	S	Y	S	G	I	V	S	G	I	E	W	A	T	T	N	G	M	D	V	I	N	M	S
Novo	S	G	Q	Y	S	W	I	I	N	G	I	E	W	A	I	A	N	N	M	D	V	I	N	M	S
			α	α	α	α	α	α	α	α	α	α	α	α	α	α	α			β	β	β	β	β	

					130										140										150
Carlsberg	L	G	G	A	S	G	S	T	A	M	K	Q	A	V	D	N	A	Y	A	R	G	V	V	V	V
Novo	L	G	G	P	S	G	S	A	A	L	K	A	A	V	D	K	A	V	A	S	G	V	V	V	V
			β				α	α	α	α	α	α	α	α	α	α	α	α	α	α	T		β	β	β
																		T	T	T					

										160										170					
Carlsberg	A	A	A	G	N	S	G	S	S	G	S	T	N	T	I	G	Y	P	A	K	Y	D	S	V	I
Novo	A	A	A	G	N	E	G	T	S	G	S	S	S	T	V	G	Y	P	G	K	Y	P	S	V	I
	β	β	β						T	T	T	T					β	T	T	T	T	T	T	T	β

					180										190										200
Carlsberg	A	V	G	A	V	D	S	N	S	N	R	A	S	F	S	S	V	G	A	E	L	E	V	M	A
Novo	A	V	G	A	V	D	S	S	N	Q	R	A	S	F	S	S	V	G	P	E	L	D	V	M	A
	β	β	β	β	β	β	T	T	T	β	β	β	T	T	T			T	T	T	T		β	β	β
						T						T													

										210										220					
Carlsberg	P	G	A	G	V	Y	S	T	Y	P	T	N	T	Y	A	T	L	N	G	T	S	M	A	S	P
Novo	P	G	V	S	I	Q	S	T	L	P	G	N	K	Y	G	A	Y	N	G	T	S	M	A	S	P
	β	β	-		β	β	β	β	β				β	β	β	β	β		T	T	T	T	T	α	α
																								T	T

					230										240										250
Carlsberg	H	V	A	G	A	A	A	L	I	L	S	K	H	P	N	L	S	A	S	Q	V	R	N	R	L
Novo	H	V	A	G	A	A	A	L	I	L	S	K	H	P	N	W	T	N	T	Q	V	R	S	S	L
	α	α	α	α	α	α	α	α	α	α	α	α	α	T	T	T	α	α	α	α	α	α	α	α	α
	T	T											T												

										260										270					275
Carlsberg	S	S	T	A	T	Y	L	G	S	S	F	Y	Y	G	K	G	L	I	N	V	E	A	A	A	Q
Novo	E	N	T	T	T	K	L	G	D	S	F	Y	Y	G	K	G	L	I	N	V	Q	A	A	A	Q
	α	α			β				α	α	α	α	α	α	T	T	β	β	α	α	α	α	α	α	T
									T	T	T	T	T	T		β						T	T	T	

Figure 2.1. Amino Acid Sequences of Subtilisins Novo and Carlsberg. Secondary structural elements, determined as described in section 2.2.1, are indicated by α (α-helix), β (β-sheet) and T (3_{10} turn). The elements indicated here are those determined for subtilisin Carlsberg; those of subtilisin Novo are very similar (section 2.2.1). Some residues participate in two kinds of secondary structure e.g. donating a hydrogen bond to residue i-4 in an α-helix and receiving a hydrogen bond from residue i+3 in a turn.

(batch 120-4). The enzyme consists of a single polypeptide chain of 275 amino acids, M_r = 27583 (Olaitan *et al.*,1968; Wells *et al.*,1983) (Fig. 2.1). The intact inhibitor contains 83 amino acids, M_r = 9250 (Svendsen *et al.*, 1980a) (Fig. 2.2). The amino acid sequence of subtilisin Novo, deduced from the DNA sequence (Wells *et al.*, 1983), shows the following changes from the published protein sequence (Olaitan *et al.*, 1968): Pro56→Asn56, Asn57→Pro57, Asp61→Asn61, Ser88→Ala88, Ala89→Ser89, Asp98→Ala98, Ala99→Asp99, Ser158→Thr158, Thr159→Ser159, Gln251→Glu251. The sequence of subtilisin Carlsberg shows one change from the published protein sequence (Smith *et al.*, 1968): Asn158→Ser158. These changes are observed in the electron density maps of the two enzyme structures. There are 82 amino acid changes between the two enzymes, and one deletion at Thr55 (Novo), thus they are 70% identical.

Small crystals of the complex, shaped like rods or hexagonal plates, were grown by the hanging-drop vapor-diffusion method (Davies & Segal, 1971; McPherson, 1982). Subtilisin Novo (7.5 mg/mL) and CI-2 (2.5 mg/mL) were dissolved in distilled water and a hanging drop was prepared consisting of equal amounts of the protein solution and a buffer of 1.4 M $(NH_4)_2SO_4$ and 50 mM KH_2PO_4 at pH 5.5. The reservoir of the crystallization tray was filled with 1.0 mL of buffer solution. The crystals grew to their maximum size in 3 to 4 weeks. They often had one dimension measuring more than 0.2 mm, but at least one dimension was always less than 0.1 mm. In order to grow larger crystals suitable for high-resolution data collection, small crystals were placed into freshly prepared hanging drops of buffer and protein solution, and grew within days to an average size of 0.7x0.6x0.15 mm.

Analysis of precession photographs showed that the crystals diffract to a minimum *d* spacing of 2.1Å. Systematic absences were observed for reflections with h+k = 2n+1. The lattice symmetry and systematic absences are consistent with monoclinic space group C2. The crystals have unit cell dimensions of *a* = 103.2(1)Å, *b* = 56.83(4)Å, *c* = 68.7(1)Å, and *β* = 127.4(2)°. The volume per unit molecular weight, V_m, is 2.20Å³/dalton, which is consistent with one molecule of the complex (M_r = 36833) per asymmetric unit

									10										20					
Eglin														T	E	F	G	S	E	L	K	S	F	P
CI-2	(Z,	V,	S,	S)	K	K	P	E	G	V	N	T	G	A	G	D	R	H	N	L	K	T	E	W	P
																•						β	β	β	T
																								T	

					30										40										50
Eglin	E	V	V	G	K	T	V	D	Q	A	R	E	Y	F	T	L	H	Y	P	[illegible]	Y	D	V	Y	F
CI-2	E	L	V	G	K	S	V	E	E	A	K	K	V	I	L	Q	D	K	P	E	A	Q	I	I	V
	T	T	T	T	T	β	β	α	α	α	α	α	α	α	α	α	α	α	T	T	T	β	β	β	β
					β	α	α											T							

										60										70			72A		
Eglin	L	P	E	G	S	P	V	T	L	D	L	R	Y	N	R	V	R	V	F	Y	N	P	G	T	N
CI-2	L	P	V	G	T	I	V	T	M	E	Y	R	I	D	R	V	R	L	F	V	D	K	-	L	D
	β	β	β	T	T							T	T	T	T	β	β	β	β	β			.		β
		T	T											β	β										

						80			
Eglin	V	V	N	H	V	P	H	V	G
CI-2	M	I	A	Q	V	P	R	V	G
	β	β				β	β	β	

Figure 2.2. Amino Acid Sequences of CI-2 and Eglin-c. The amino acid sequence of CI-2 was determined by Svendsen *et al*. (1980a), that of eglin-c by Seemüller *et al*. (1980). Notation as for Figure 2.1. The secondary structural elements indicated here are those for eglin-c; those of CI-2 are very similar (section 2.3). CI-2 is subject to hydrolysis during purification between Asn11I and Thr12I, Gly15I and Asp16I, and Arg17I and His18I (Svendsen *et al*., 1980a).

and a solvent content of 44% (Matthews, 1968).

2.1.2 Data Collection and Processing

Intensity data sets were collected from two crystals, described in Table 2.1. Both data sets were collected on a Nonius CAD4 diffractometer. The incident radiation was Ni-filtered CuK_α, at 40 kV and 26 mA, using a 1.3 mm diameter incident beam collimator. The crystal-to-counter distance was 60 cm, and the diffracted beam passed through a He-filled tunnel. The peaks were scanned 0.5° in ω at 0.67°/min., and backgrounds were scanned 0.125° in ω on either side of the peak scan. To minimize loss of weak reflections at higher resolution due to decay, the data for the 2.1Å data set were collected in three shells; first from 2.1 to 3.0Å, then from 3.0 to 5.0Å, and finally from 5.0 to 43.9Å.

During data processing, the measured backgrounds were corrected for intensity dependence and an averaging function of 2θ and ϕ was applied. The absorption correction applied was an empirical function of ϕ (North *et al.*, 1968). Lorentz and polarization corrections were also applied. The absolute scale and an overall temperature factor (B, Å^2) were calculated with the program ORESTES (Thiessen & Levy, 1973).

Intensity decay due to radiation damage was evaluated by collecting the same set of approximately 1000 reflections at the beginning and end of data collection and analyzing

Table 2.1

Crystal Data for the Subtilisin Novo:CI-2 Complex

	Crystal 1	Crystal 2
Resolution range (Å)	43.5 - 2.5	43.9 - 2.1
Total reflections measured	13174	21672
Total unique reflections	11071	18705
Total reflections $I \geq \sigma(I)$[1]	9695	16539
R_{merge}[2]	0.037	0.064
Maximum absorption correction	2.55	2.09
Maximum decay correction	7 %	11 %
Scale	33.44	28.99
Overall B (Å^2)	24	15

[1] $\sigma(I) = (I + c^2 I^2 + (t_I/t_{Bk})^2 (\Sigma Bk + c^2 \Sigma Bk^2))^{1/2}$, where I = total intensity; Bk = background counts; t_I = time for intensity measurement; t_{Bk} = time for background measurement; c = instrument instability constant = 0.01.

[2] $R_{merge} = \Sigma_{hkl}(\Sigma_i |I_i - \langle I \rangle| / \Sigma_i I_i)$ for reflections measured more than once in a data set.

them for decay as a function of time and 2θ by the method of Hendrickson (1976). Since the diffractometer was out of order for two weeks in the middle of the 2.1Å data collection, a third set of decay standards was collected when it was restarted. The overall decay was analyzed on the assumption that decay as a function of time took place at the same rate when the diffractometer was not working as when data were being collected. Five standard reflections were also collected periodically, and the adequacy of the decay correction was tested with them. The intensity of each of the five was corrected as a function of time, first without including the two weeks of diffractometer malfunction, and then with the assumption above. The corrected intensities were then plotted against time, omitting the lost two weeks. The two plots were very similar, and no large discontinuity was seen in the first plot at the point when the diffractometer was down. This indicates that either method would have provided adequate decay correction for this data set, but this may not be true for data sets with greater amounts of overall decay. In this data set, the low amount of overall decay makes the detection of differences from the two methods difficult, but also reduces the need for a highly accurate decay correction.

2.1.3 Structure Determination

The method of molecular replacement (Rossmann, 1973) was used to solve the phase problem for the structure of CI-2 in complex with subtilisin Novo. The search model was the crystal structure of subtilisin BPN', with atomic coordinates for the model taken from the Brookhaven Protein Data Bank. These coordinates had been partially refined by difference Fourier methods to an *R*-factor of 0.44 at 2.5Å resolution (Alden *et al.*, 1971). This was considered to be a good search model for two reasons; first, the model enzyme has the same sequence as the enzyme in the complex, and was believed to have a very similar overall conformation, although the two might differ in details; second, the model accounts for 77% of the scattering matter expected in the complex, a relatively high percentage. Data from crystal 1 (Table 2.1) were used for the molecular replacement calculations.

The first step in the molecular replacement was to find the orientation of the search model in the unit cell of the complex. The rotation function search was performed with the fast rotation function of Crowther (1973), using a program written by E. Dodson and modified locally by L. Sawyer and R. Read. The rotation function attempts to maximize the agreement between the Patterson maps of the model and the unknown structure (Rossmann & Blow, 1961). The Patterson function (Patterson, 1934) is the Fourier transform of the intensities (*I*) or the squares of the structure factors ($|\mathbf{F}|^2$). The map of this function contains peaks at positions corresponding to all the interatomic vectors in the structure. It includes both intra- and inter-molecular vectors from within the crystal. Since the shortest interatomic vectors will be mainly intra-molecular vectors, the majority of peaks close to the origin of the Patterson function will be those arising from these intramolecular vectors.

The vectors for the model structure will have a particular orientation relative to the origin of the Patterson map. If the model and the unknown structure are similar, their structure factors will give rise to a similar set of interatomic vectors, but the two sets will probably not be oriented in the same direction relative to the Patterson origin. The fast rotation function search is equivalent to rotating the Patterson map of the model structure relative to that of the unknown structure and checking for the orientation that maximizes the match between the two. The search is done only in the area of the maps close to the origin. This excludes most inter-molecular vectors from the problem, since they will be different for the model and unknown structures even when the intra-molecular vectors have been matched correctly.

The coefficients of the rotation function search for the subtilisin Novo:CI-2 complex were $|\mathbf{E}|$'s (normalized structure factors). The $|\mathbf{E}_c|$'s were calculated for the subtilisin BPN' model placed in an orthogonal unit cell of symmetry P1, with each edge of the cell measuring 65Å. The cell dimensions are calculated to be 10 to 15Å larger than the diameter of the model molecule, to minimize the number of inter-molecular vectors close to the origin of the Patterson map (Lifchitz, 1983). The $|\mathbf{E}_o|$'s were calculated for the unknown structure, from the structure factors derived from data processing, with the program ORESTES (Thiessen & Levy, 1973). The radius of integration for the search was 4 to 18Å (the Patterson maps are only compared between these distances from the origin, to limit the comparison to intra-molecular vectors). The resolution range of reflections included in the search was 10.0 to 3.0Å. The limit of 3.0Å was chosen to be of medium resolution; past experience in this lab has shown that using a higher resolution limit does not work as well because the details of the model and unknown structures tend to differ at high resolution; using a lower resolution limit leads to broader peaks in the rotation function (Read, 1986a). The limit of 10.0Å omits very low resolution reflections that primarily contain information about overall molecular shape and bulk solvent. The number of $|\mathbf{E}|$'s used in the rotation function calculation is limited by the capacity of the program to 2500. In addition to the resolution limits, a minimum size cutoff was used to restrict the data chosen. A minimum acceptable value for $|\mathbf{E}|$ of 2.5 gave 2368 $|\mathbf{E}_c|$'s. The cutoff for $|\mathbf{E}_o|$ was set at 1.0, which gave 1864 reflections, close to the number of $|\mathbf{E}_c|$'s. Past experience shows that maximizing the number of reflections accepted increases the height and accuracy of the rotation function solution (Read, 1986a).

Two rotation function searches were performed, one with subtilisin BPN' as the model, described above, and one with the 2.8Å subtilisin Novo coordinates from the Brookhaven Protein Data Bank. The initial coarse search was performed on an asymmetric unit of rotation function space (Rao *et al.*, 1980), along a grid of 5° in each of the Eulerian angles α, β, and γ (Crowther, 1973). The maximum peak for the Novo model in this search was 10.5σ above the mean and 6.1σ above the next highest peak; that for the BPN' model was 12.7σ above the mean and 8.5σ above the next highest peak. Both these results indicate the correct solution to the rotation problem quite unambiguously; the section of the rotation function map containing the peak for the BPN' model is shown in Figure 2.3. The BPN' model may have given a higher peak because the data extend to higher resolution and have been partially refined. The BPN' model was used for all subsequent steps in the molecular replacement structure solution.

Another rotation search was performed at 1° intervals in β and 5° intervals in α and γ about the maximum peak from the coarse rotation search. This fine search gave a maximum peak of 11.2σ above the mean of that map. The final values of the orientation parameters from the rotation search were

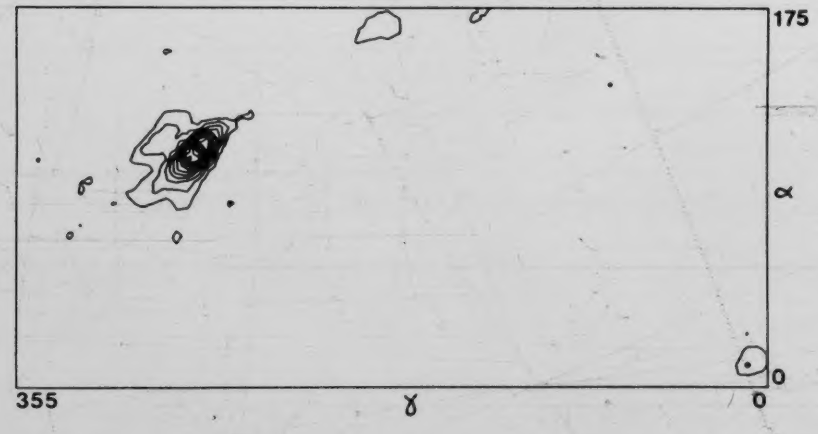


Figure 2.3. Rotation Function Map, Subtilisin Novo:CI-2 Complex. The section of the rotation function map, from the molecular replacement solution of the subtilisin Novo:CI-2 complex structure, containing the maximum peak. The Euler angle β is 50° in this section. The model for this search was subtilisin BPN'. Contours are in increments of 1σ from the mean value of the map and begin at 2σ.

interpolated from the fine search map to 0.2° in β and 1° in α and γ. The program does not allow a finer search in α and γ due to considerations of computing time.

Once the correct orientation of the unknown structure in the unit cell has been found, it must be placed correctly relative to the origin and symmetry elements. This is the translation problem. The translation problem for the subtilisin Novo:CI-2 complex was solved using a program written locally by M. Fujinaga, called BRUTE. The principle of this search is similar to that of the rotation function; structure factors are calculated for the rotationally oriented model at positions on a grid throughout the unit cell, and these are compared to the structure factors of the unknown molecule. The algorithm used is a correlation coefficient (r) search on $|\mathbf{F}|^2$, where

$$r=\Sigma(x-\overline{x})(y-\overline{y})/[\Sigma(x-\overline{x})^2\Sigma(y-\overline{y})^2]^{1/2}$$

In this algorithm, $|\mathbf{F}_O|^2 = x$ and $|\mathbf{F}_C|^2 = y$. The correlation coefficient search is preferable to the more usual *R*-factor search because r is insensitive to scaling errors and gives smoother and more consistent maps (M. Fujinaga, personal communication).

The translation function search must be done over a volume of the unit cell that contains a single permissible origin; this volume is not necessarily the same as the asymmetric unit of the unit cell. For the space group C2, all points on the *b*-axis are permissible origins, as are the axes (1/2, y, 0) and (0, y, 1/2). The translation search area is thus a plane perpendicular to the *b*-axis with $0 \le x \le 1/2$, $0 \le z \le 1/2$. Due to limitations in the capacity of computing equipment, the search program can accept only 4000 reflections. Limiting the resolution range is the method used to limit the number of reflections; data between 4.0 and 5.0Å were used in this search. Past experience in this lab was the basis for this choice of resolution range (Read, 1986a). The use of relatively low resolution data minimizes the effect of detailed differences between the model and the unknown structure, as well as minimizing the number of grid points that must be evaluated in the search. The 5.0Å limit excludes low resolution data that contains information about disordered solvent; since this solvent is not included in the molecular replacement model, low resolution data will not contribute significantly to the solution. In the initial coarse translation search, the grid spacing over the search area was set at $1/4d_{min}$, or 1.0Å. This grid spacing is again a compromise between the desire to examine a very fine grid to ensure that no peaks are missed between grid points, and limitations on available computing power. The rule of thumb that the grid should be $1/4d_{min}$ derives from trials and numerical experiments (M. Fujinaga, personal communication). In general, the translation search should be repeated with the translation origin offset 1/2 grid unit along each axis, in order to effectively decrease the size of the grid. With a grid of $1/4d_{min}$, in a single search only the shoulder of a peak may be seen that may not be distinguishable from noise. The height of the peak found by the first search in this structure solution made a second search unnecessary.

With the model coordinates rotated to the orientation found with the rotation function search, a translation search with the above parameters gave a peak of 8.0σ above the mean and 2.5σ above the next highest peak. This peak in the plane searched by the translation function is shown in Figure 2.4. A second translation function search was made

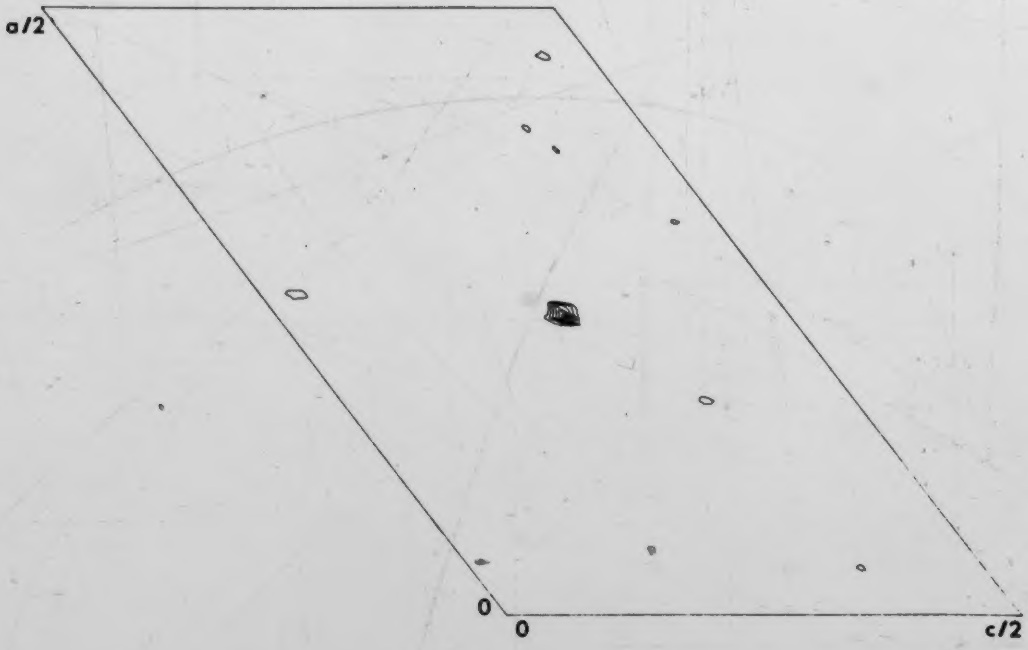


Figure 2.4. Translation Function Map, Subtilisin Novo:CI-2 Complex. The section of the translation function map, from the molecular replacement solution of the subtilisin Novo:CI-2 complex structure, containing the maximum peak. The xz plane was searched at y = 0, x = 0 → 1/2, z = 0 → 1/2. Contours are in increments of 1σ from the mean value of the map and begin at 2σ.

about this peak at grid intervals of 0.1Å, extending ±1.0Å along the x and z axes from the peak. The maximum of the fine search was 9.0σ above the mean of the coarse search map. The BRUTE program also allows refinement of the rotational parameters of the solution, using a 6-dimensional rotation and translation search over a small area of the map. Although this option has improved significantly the rotational parameters for other structures solved in this lab, in this case the orientation found by BRUTE in a 6D search about the translation function peak was identical to that found by the rotation function search alone. The unambiguity and sharpness of the rotation and translation function peaks for this problem are probably due to the high degree of resemblance between the model and the unknown structure.

The R-factor for the subtilisin BPN' model, rotated and translated by the parameters of the molecular replacement solution, was 0.52 for all data to 2.5Å resolution. This model, with no contribution from any model for the inhibitor, was used in the calculation of structure factor amplitudes and phases for the computation of an initial electron density map. The coefficients for all maps calculated in this structure solution were derived from an expression designed to suppress model bias resulting from phasing by partial structures with errors (Read, 1986b). Unrefined or partially refined models of macromolecules are generally incomplete and typically have large coordinate errors. Electron density maps calculated with phases from models are biased toward the model, so Fourier coefficients that reduce the amount of model bias will give electron density maps that more closely resemble the true structure. The coefficients suggested by Read (1986b) are

$$F_N \simeq (2m|F_N|-D|F_P^c|)\exp(i\alpha_P^c)$$

where F_N is the structure factor for the complete structure, m is the expected value of the cosine of the phase error (= figure of merit), D is the coefficient to reduce model bias, F_P^c is the structure factor of the partial structure with errors, and α_P^c is the phase of the partial structure with errors.

$$D = \langle\cos(2\pi \mathbf{s}\cdot\Delta\mathbf{r})\rangle$$

(Luzzati, 1952), thus if there are no coordinate errors D = 1.

2.1.4 Structure Fitting and Refinement

Throughout the fitting and refinement of the subtilisin Novo:CI-2 complex structure, the MMS-X interactive graphics (Barry *et al.*, 1976) with the macromolecular modelling system M3, developed by C. Broughton (Sielecki *et al.*, 1982), was used for map interpretation and model fitting.

The backbone of most of the subtilisin Novo molecule, as well as many side-chains, could be seen easily in the initial electron density map calculated from the subtilisin

BPN' model. The density for the chain around Ser98 to Ser104 and Gly154 to Thr164 was weak and somewhat discontinuous, and short breaks in the main-chain occurred at several other points, but the entire course of the chain could still be followed readily with the model. This was one confirmation of the assumption made during the molecular replacement procedure that the conformation of the enzyme in the complex did not change significantly from that of the native enzyme. Before refinement of the complex was begun, the entire subtilisin molecule was refitted to the electron density map, although most of the refitting consisted of only minor adjustments to main-chain and side-chain positions.

As a starting model for the inhibitor CI-2, the P_4 to P_3' residues (the reactive site loop)[3] from the third domain of the ovomucoid inhibitor from turkey eggs (OMTKY3) was fitted to electron density in the active site of the enzyme. The structure of this inhibitor has been solved in complex with a bacterial serine proteinase, *Streptomyces griseus* proteinase B (SGPB) (Fujinaga *et al*., 1982; Read *et al*., 1983), and with α-chymotrypsin (Read *et al*., 1984). The side-chains of these 7 residues were 'mutated' to the sequence at the reactive site loop of CI-2, retaining the coordinates of the main-chain atoms and the appropriate side-chain torsion angles from the known structure (program of R. Read). The conformations of the two inhibitor reactive site loops were expected to be similar, because all inhibitors acting by the standard mechanism of Laskowski & Kato (1980) are thought to bind to the enzyme in the manner of a good substrate. In fact, the mutated OMTKY3 model fit the density in the active site of subtilisin Novo very well.

---

[3] The notation of Schechter and Berger (1967) is used to facilitate discussion of the interactions between a proteinase and bound peptides. Amino acid residues of substrates are numbered P_1, P_2, etc., in the N-terminal direction and P_1', P_2', etc. in the C-terminal direction from the scissile bond. The complementary subsites of the enzyme binding region are numbered S_1, S_2 and S_1', S_2', etc.

A 'mini-map' of the electron density contours for the complex was calculated and copied onto clear plastic sheets that could be stacked up to provide a three-dimensional view of the map. This map was used in conjunction with information on the packing of the enzyme in the unit cell to determine a molecular boundary for the inhibitor. Electron density was present throughout the volume believed to be occupied by the inhibitor, but was weak in many places. Short stretches of polypeptide backbone could be identified both in the 'mini-map' and on the MMS-X graphics system, but the connectivity between them was poor. Residues from Val53I to Tyr61I , plus two segments of four residues each of poly-L-alanine that appeared to be part of a β-sheet structure close to the reactive site loop, were fitted into density. This was the extent of the chain that could be fitted with some certainty, and these three segments were all of the inhibitor that was included at the beginning of the refinement.

The refinement of the structure of the complex was carried out with the restrained-parameter least-squares refinement program of W. Hendrickson and J. Konnert (1980), modified locally by M. Fujinaga for the FPS164 attached processor. The purpose of refinement is to improve the agreement between the observed structure factors (F_o) and those calculated from the atomic coordinates of the current model (F_c), while maintaining good stereochemistry in the protein structure. Since much is known about the geometry of amino acids and peptide bonds from small molecule crystal structures, this information can be included in the refinement as restraints on stereochemistry. In this refinement, restraints were placed on covalent bond lengths, interbond angles, planar groups (peptide bonds, aromatic rings, guanidinium groups, carboxyl groups, amide groups, imidazole groups), chirality of chiral centers, non-bonded contacts, hydrogen-bond distances, and ω torsion angles. Restraints were also applied to individual atomic temperature factors when they were added to the refinement. The practical application of restraints in protein structure refinement has been discussed by Hendrickson (1981).

The course of this refinement is summarized in Table 2.2 and the changes in the *R*-factor during the refinement are shown in Figure 2.5. Data from crystal 2 (Table 2.1) were used for the entire refinement. The general strategy employed in the refinement was to fit the model to an electron density map as well as possible, omitting residues with

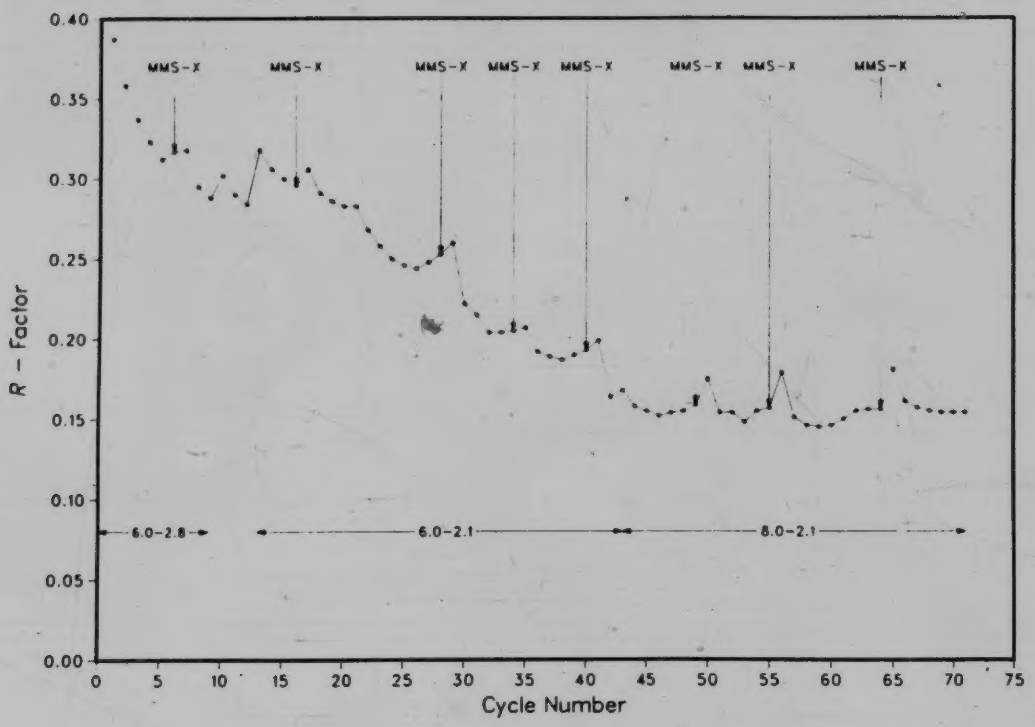


Figure 2.5. Progress of the Subtilisin Novo:CI-2 Complex Refinement. The *R*-factor at each cycle is plotted versus cycle number. The resolution range of the data included at each cycle is indicated. Data from 6.0 to 2.5Å were used for cycles 10 to 13. Electron density maps with coefficients $2m|F_o|-D|F_c|$, α_c were calculated and used to refit the model on the graphics system at the points marked "MMS-X".

very weak or uninterpretable density. This model was refined for several cycles, starting with fairly loose geometric restraints (e.g. 0.05Å on bond lengths, 0.06Å on interbond angles). As the refinement began to converge, the restraints were tightened sufficiently to produce a model with good stereochemistry for the next examination on the graphics system (e.g. 0.02Å on bond lengths, 0.03Å on interbond angles). Structure factors were calculated from

Table 2.2

Course of Least-squares Refinement

Cycle number	Data used(Å)	Number of reflections	R-factor	Comments
1	6.0-2.8	6662	0.387	Start refinement with model containing entire subtilisin Novo sequence[1], residues 53I to 61I + 2 segments of 4 alanines each for inhibitor. Use all reflections with $I \geq \sigma(I)$. One overall *B*-factor.
6	6.0-2.8	6662	0.317	Refit enzyme and inhibitor, add Pro52I, Arg62I to Arg65I, and 14 more alanines in inhibitor density.
7	6.0-2.8	6662	0.318	Restart with 28 inhibitor residues.
10	6.0-2.5	9422	0.302	Add in data from 2.8 to 2.5Å.
13	6.0-2.1	15661	0.318	Add in data from 2.5 to 2.1Å.
16	6.0-2.1	15661	0.296	Refit inhibitor, adding Val28I-Leu51I and removing alanines placed in disconnected density.
17	6.0-2.1	15661	0.306	Restart with 38 inhibitor residues.
21	6.0-2.1	15661	0.283	Start refining with individual atomic *B*-factors.
28	6.0-2.1	15661	0.253	Refit enzyme, using corrected sequence from DNA[2]. Add Lys21I to Leu27I, Val66I to Gly83I to inhibitor, refit remainder of inhibitor.
29	6.0-2.1	15661	0.260	Restart with 63 inhibitor residues.
34	6.0-2.1	15661	0.205	Adjust enzyme and inhibitor models in new electron density map.
35	6.0-2.1	15661	0.207	Restart refinement.
40	6.0-2.1	15661	0.193	Adjust enzyme and inhibitor models. Calculate difference map, coefficients $m\|F_o\|-\|F_c\|$, α_c. Choose solvent positions from difference map peaks. Add Leu20I to inhibitor.
41	6.0-2.1	15133	0.199	Restart with 107 H_2O and 1 Ca^{2+} in model, 64 inhibitor residues. Ca^{2+} refined as sulphur atom.
43	8.0-2.1	16128	0.168	Add data from 8.0 to 6.0Å.

(Table 2.2 continued)

49	8.0-2.1	16128	0.159	Adjust model. Remove H_2O with high *B*-factors and low occupancies. Choose new solvent from difference map.
50	8.0-2.1	16128	0.175	Restart with 94 H_2O and 2 ions in model.
55	8.0-2.1	16128	0.157	Remove poor H_2O, choose new ones from difference map. Adjust model.
56	8.0-2.1	16128	0.179	Restart with 141 H_2O and 2 ions in model.
64	8.0-2.1	16128	0.156	Adjust model in minor ways. Choose new H_2O molecules.
65	8.0-2.1	16128	0.181	Restart with model including 165 H_2O. Refine two ions as Ca^{2+}.
71	8.0-2.1	16128	0.154	End of refinement.

---

[1] Olaitan *et al.*, 1968.
[2] Wells *et al.*, 1983.

the new model and a new electron density map calculated with the coefficients described in section 2.1.3. The model was manually adjusted to fit the new map and residues added in improved density where possible. At later stages of the refinement, difference electron density maps with coefficients $m|\mathbf{F}_o|-|\mathbf{F}_c|, \alpha_c$ were also calculated at each refitting. These maps were searched for significant peaks, using a local program, and the peaks examined on the graphics system for indications of errors in fitting the model or for the presence of solvent molecules.

Following this strategy, a model for the inhibitor CI-2 was gradually built. The tracing of the chain of the inhibitor was essentially complete by cycle 29; with refinement of the model, the connectivity of the inhibitor chain became much more clear, and alanine residues that had been placed in stretches of disconnected density were replaced with the proper sequence of the inhibitor. The model extends only to Leu20I at the N-terminus of CI-2; electron density is lacking for the first 19 residues. The regions of weak density that gave rise to the discontinuities in the chain in the early maps are loops connecting elements of secondary structure in the inhibitor. These loops are located at the edges of the molecule and have higher than average temperature factors in the fully refined structure.

Because the molecular replacement model for subtilisin Novo was so similar to the structure of the enzyme in the complex, the manual adjustment of the model was not extensive, even in the early stages of refinement. At cycle 40, a new sequence of subtilisin Novo became available, derived from the DNA sequence (Wells *et al.*, 1983). The DNA sequence contained the following changes from the published protein sequence (Olaitan *et al.*, 1968): Pro56→Asn56, Asn57→Pro57, Asp61→Asn61, Ser88→Ala88, Ala89→Ser89, Asp→Ala98, Ala99→Asp99, Ser158→Thr138, Thr159→Ser159, Gln251→Glu251. Examination of the electron density map at that stage showed unambiguously that the DNA sequence was correct in six of these changes. The inversion of Ser89 and Ala 88 was less clear in the map, but the change was made in the model and the two residues subsequently refined well in their new positions. The difference between Asp and Asn or Glu and Gln usually cannot be detected directly by X-ray crystallography, but Asp61 was changed to Asn61 and Gln251 to Glu251 for the rest of the refinement.

Solvent molecules were chosen from difference electron density maps (coefficients decribed above), beginning at cycle 40. It is often difficult to distinguish real solvent molecules from noise in a difference map; real atoms that are not included in the model used to phase the map appear at only 1/2 their weight (Luzzati, 1953). In addition, solvent molecules are often weakly bound to the protein and appear as weaker peaks on a difference map because their occupancy of a position is less than 1.0. To minimize the inclusion of noise peaks as solvent molecules, only strong peaks with well-defined convex shapes in the electron density map were chosen from each difference map. Solvent molecules included in the model were refined with both variable *B*-factors and occupancies. If the *B*-factor of a solvent molecule rose above 26Å^2, or its occupancy dropped below 0.40, it was considered to be noise and was removed from the model. Because of these fairly stringent criteria, the final model includes only 167 solvent molecules. 165 of these are interpreted as water molecules; the remaining 2 are believed to be Ca^{2+} ions, based on their refined occupancies and *B*-factors, their coordination geometry, and their characteristic non-bonded contact distances to ligands. When treated as water molecules, the occupancies of the 2 ions refined rapidly to 1.0, and their *B*-factors to the lowest value allowed by the program. They were subsequently refined as sulphur atoms to cycle 64, when the scattering factors for Ca^{2+} ions were added to the information provided to the refinement program. The water molecules from the final refinement cycle were ordered according to an empirical 'quality factor', defined as occupancy^2/*B* (James & Sielecki, 1983). Thus, water molecules with low sequence numbers are relatively more reliable than those with high sequence numbers (sequence numbers of water molecules in this structure begin at 342).

2.1.5 Quality of the Refined Complex Structure

The completeness of a refinement and the quality of the resulting crystal structure can be assessed by a number of criteria. A refinement can be considered complete if the model has essentially stopped changing, and there are few indications that further changes in the model are necessary. The quality of the structure can be judged by measures of the accuracy of coordinates, the reasonableness of stereochemistry, and the agreement between observed and calculated structure factors.

The refinement of the subtilisin Novo:CI-2 complex was considered complete after 71 cycles. The final refinement parameters are summarized in Table 2.3. The indicated parameter shifts for this final cycle are very small. The root-mean-square (rms) coordinate shift is 0.012Å; the maximum coordinate shifts are 0.082Å for atoms in the enzyme, 0.075Å for atoms in the inhibitor, and 0.230Å for water molecules. The largest shifts in the proteins are for atoms in lysine side-chains. The majority of coordinate shifts for both protein and solvent atoms are less than 0.010Å.

A difference electron density map was computed at cycle 71 and examined on the graphics system. The highest peak on

Table 2.3

Final Refinement Parameters and Results

No. of cycles	71
R-factor	0.154
Resolution range (Å)	8.0-2.1
No. of reflections ($I \geq \sigma(I)$)	16128
No. of protein atoms	2451
No. of solvent atoms[1]	167
No. of variable parameters	10640
$<\|\|F_o\|-\|F_c\|\|>$	67
<Coordinate shift> (Å)	0.012
<*B*-factor shift> (Å^2)	0.25
Rms deviations from ideal values[2]	
Distance restraints (Å)	
Bond distance	0.006(0.008)
Angle distance	0.024(0.016)
Planar 1-4 distance	0.018(0.016)
Plane restraint (Å)	0.013(0.012)
Chiral-center restraint (Å^3)	0.119(0.080)
Non-bonded contact restraints (Å)	
Single torsion contact	0.276(0.400)
Multiple torsion contact	0.202(0.400)
Possible hydrogen bond	0.183(0.400)
Isotropic thermal factor restraints (Å^2)	
Main-chain bond	1.721(2.000)
Main-chain angle	2.258(2.000)
Side-chain bond	3.866(3.000)
Side-chain angle	5.183(3.000)

[1] Including 2 Ca^{2+} ions.

[2] The values of σ, in parentheses, are the input estimated standard deviations that determine the relative weights of the corresponding restraints [see Hendrickson and Konnert (1980)].

the difference map was 0.79e/Å^3.[4] The map contained 7 additional peaks with heights of 21 to 24e/Å^3 and 11 negative troughs with depths in the same range. The highest peak is superposed on Ca^{2+} ion 1. The small parameter shifts for this ion in the final cycle of refinement indicate that its positional parameters are accurate, thus its *B*-factor is probably overestimated. The height of this residual peak in the difference map is small compared to the total peak height expected for a Ca^{2+} ion.

---

[4] Electron density maps were calculated omitting the contribution of the F_{000} term and thus have a mean electron density of zero. Contour levels therefore do not refer to the actual electron density but rather to the deviation above the mean density of the map. Contour levels for difference ma[illegible] refer to actual electron density because the F_{000} term i[illegible]on to calculated and observed structure factors.

Six of the remaining 7 positive peaks are most likely additional solvent molecules. The 7th may be a weakly occupied alternate conformation for the side-chain of Glu26I. Nine of the 11 negative troughs are associated with segments of the model that are not properly centered in the electron density; the density does not conform to a model with good geometry. All of these deviations are small. One trough is 0.6Å from the $O^{\delta 1}$ of Asp259, a residue with poor side-chain density and *B*-factors of 25 to 42Å², that could be somewhat misplaced. The final trough is 0.6Å from the C^{ϵ} of Met59I, the P_1 residue of the inhibitor. The density at the tip of this side-chain is a round blob surrounding the S^{γ}, with little indication of the C^{ϵ} position; its current position in the model may be wrong.

An estimate of the accuracy of the coordinates is particularly important for crystal structures of enzymes, where the catalytic mechanism is of special interest. Small changes in distances between important functional groups, on the order of 0.1Å, can affect the validity of proposed chemical reaction schemes. Unfortunately, estimated standard deviations for atomic parameters are not available from the current version of the refinement program, thus coordinate errors must be estimated by less direct means. One commonly used method for protein structures is that of Luzzatti (1952). The variation of the *R*-factor as a function of sin θ/λ is compared to a theoretical variation to determine an overall coordinate error for the molecule. The main assumption of this method is that differences between the observed and theoretical variations are due only to coordinate errors. Differences due to errors in data measurements and errors in the protein model from omission of hydrogen atoms and disordered solvent, that can be an important part of the observed differences, are not taken into account.

For the subtilisin Novo:CI-2 complex, coordinate errors were estimated by two methods: the σ_A plot of Read (1986a,b), and the method of Cruickshank (1949, 1954, 1967). The σ_A plot gives an overall estimate of the coordinate errors in the structure based on a derivation similar to that of Luzzati, but with a somewhat different set of basic assumptions. Omission of some atoms from the model is accounted for, although the assumption is made that the missing atoms are of the same type and have the same overall *B*-factor as the atoms of the model. Errors in data measurement can still be a significant part of the estimated coordinate errors, especially for a highly refined structure. The σ_A plot is easier to interpret than the standard Luzzati plot because the coordinate errors are derived from the

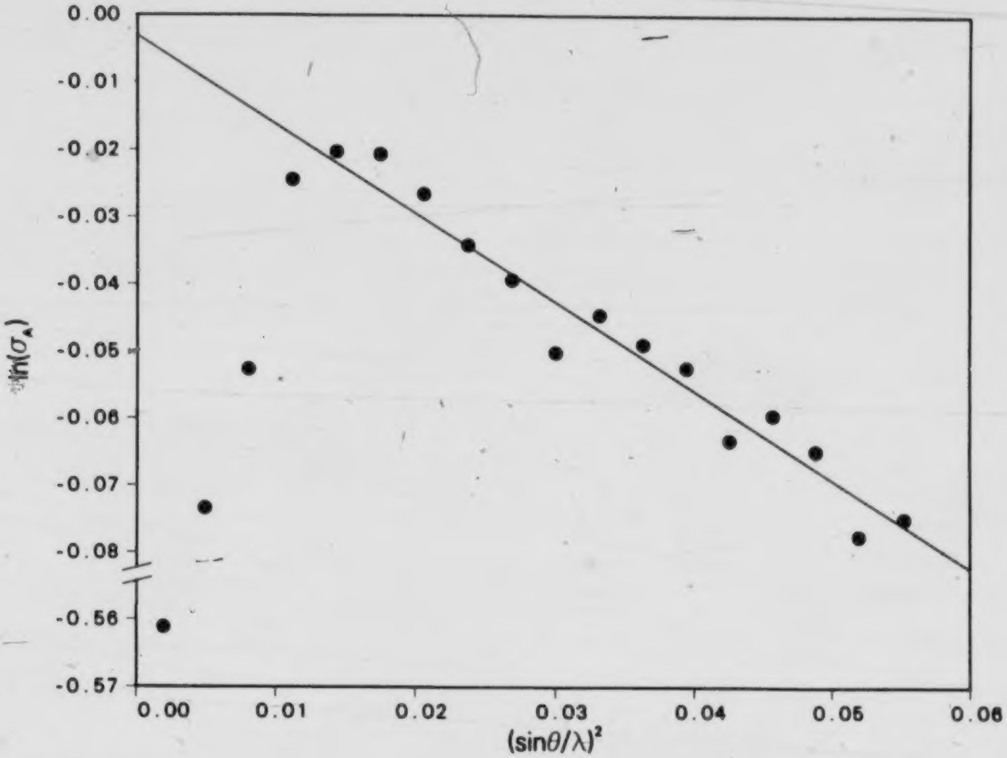


Figure 2.6. σ_A Plot to Estimate Coordinate Error. The slope of the straight-line portion of this plot is $[-\pi^3(<|\Delta r|>)^2]$ (mean coordinate error), or $[-(8\pi^2/3)<|\Delta r|^2>]$ (rms coordinate error). The line is determined by a least-squares fit to all data points but the first three. The first three data points describe low resolution data that is affected by omission of bulk solvent from the protein model.

slope of a straight line, rather than comparison with a family of standard curves. The σ_A plot is also unaffected by errors in scaling the structure factor data, since it is calculated with normalized structure factors.

The σ_A plot for the fully refined structure of the subtilisin Novo:CI-2 complex is presented in Figure 2.6. The slope of the straight-line portion of this plot is -1.318Å² and the corresponding mean coordinate error is 0.206Å. The rms coordinate error is 0.224Å. These estimated errors are larger than those observed for other refined protein structures; for proteins of a similar size, refined to similar *R*-factors, the errors are in the range of 0.10 to 0.15Å (James & Sielecki, 1983; Read *et al.*, 1983). The accuracy of atomic coordinates, however, is limited by the resolution of the data available. Crystals of the subtilisin Novo:CI-2 complex are more disordered than those of the above proteins, and do not diffract to as high a resolution limit (2.1Å vs 1.7 to 1.8Å). This disorder leads to less certainty in the determination of atomic positions, and thus larger estimated coordinate errors.

An estimate of individual atomic coordinate errors is often more meaningful than the overall estimate obtained from the σ_A plot or the method of Luzzati. A modification of the formulae of Cruickshank (1949, 1954, 1967) can be used to obtain an estimate of the coordinate error for each atom type as a function of *B*-factor. Cruickshank's formulae originally were derived for the determination of errors in peak positions in electron density maps, not for errors in atomic coordinates. Peaks in maps are assumed to correspond to atomic positions, and atoms are assumed to be well-resolved from each other. This second assumption is not generally true for protein structures. Despite these theoretical problems, Cruickshank's method gives results that are very close to error estimates from other methods

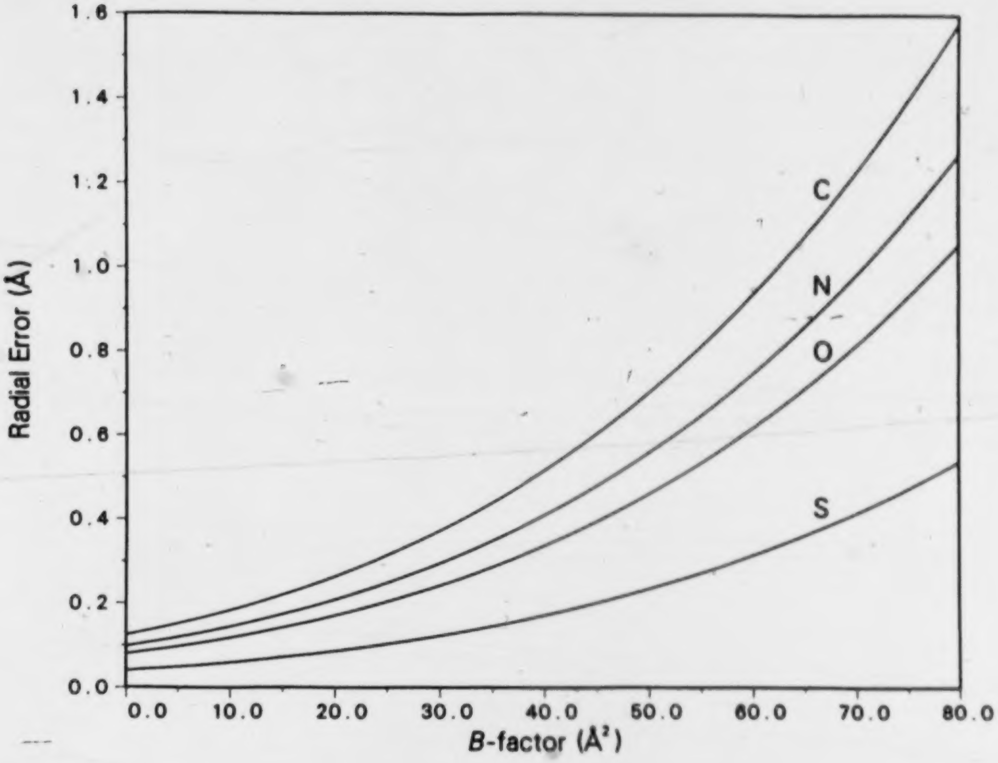


Figure 2.7. Atomic Coordinate Error by the Method of Cruickshank. The estimated radial standard deviations in atomic position are given as a function of *B*-factor. The 4 curves, from top to bottom, are for carbon, nitrogen, oxygen, and sulphur atoms in the final refined structure of subtilisin Novo and CI-2. No curve is shown for Ca^{2+}; the two Ca^{2+} ions have estimated standard deviations of 0.05Å (ion 1) and 0.11Å (ion 2).

(Read *et al.*, 1983).

Read (1986a) has shown that Cruickshank's estimated coordinate error must be divided by the number of centering translations for centered space groups; this factor is 2 for the space group C2. The resulting formula for a monoclinic space group, for the estimated standard deviation in the y coordinate for atom i, is

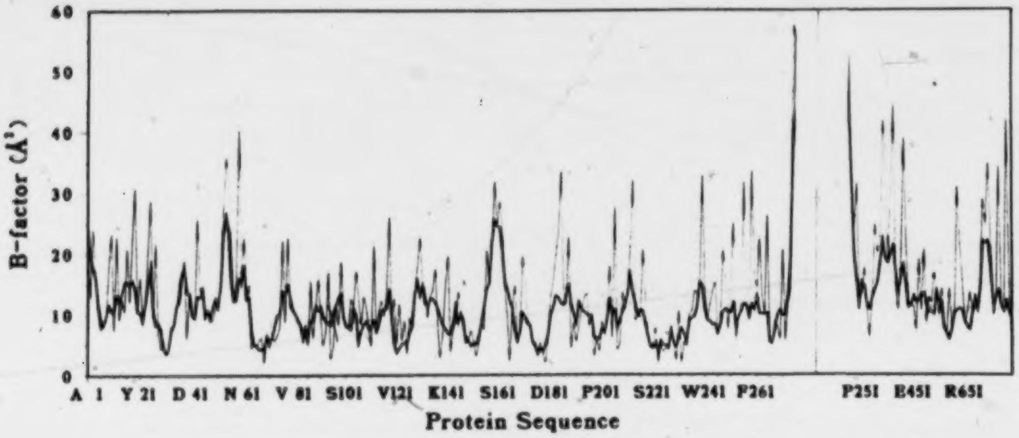


Figure 2.8. Variation in *B*-factor Along the Polypeptide Chain. The heavy line denotes the mean *B*-factor of the main-chain atoms, the light line that of the side-chain atoms. The vertical line separates enzyme and inhibitor residues. Inhibitor residue numbers are followed by an 'I' to distinguish them from those of the enzyme. No *B*-factors are given for residues 1I-19I of the inhibitor; they are not seen in the electron density map. The mean overall *B*-factor for the complex is 13Å².

$$\sigma(y_i) = \frac{b}{4\pi} \times \frac{[\Sigma k^2(|F_o|-|F_c|)^2]^{1/2}}{\Sigma (m/2) k^2 f_{0i} \exp[-B_i(\sin\theta/\lambda)^2]}$$

where b is the axial length, f_{0i} is the atomic scattering factor and $m = 2$ or 1, depending on whether the reflection is centric or non-centric, respectively. In the monoclinic case, $\sigma(x) = \sigma(z) = \sigma(y)$. The radial error in atomic position is calculated as $\sigma(r_i) = \sqrt{3}\sigma(y_i)$. Radial errors are calculated for each atom type in the structure and for a range of *B*-factors. A plot of radial error as a function of *B*-factor is given in Fig. 2.7 for the carbon, nitrogen, oxygen, and sulphur atoms in the subtilisin Novo:CI-2 complex structure. The data in this plot can be used with that in Figure 2.8, showing *B*-factors as a function of position in the polypeptide chain, to obtain an estimate of the accuracy of a specific region of the structure.

An overall estimated coordinate error can be determined from the method of Cruickshank by calculating the rms value of the estimated radial errors for all the atoms in the model. This value for the subtilisin Novo:CI-2 complex is 0.210Å, very close to the value of 0.224Å obtained from the σ_A plot.

Another measure of the quality of a crystal structure is the deviation of the geometry of the model from ideal values. Values for deviation from ideality of some geometrical parameters are given in Table 2.3; the ideal values are derived from small molecule crystal structures (Sielecki *et al.*, 1979). These deviations for the subtilisin Novo:CI-2 complex are in the range of the deviations found for the small molecule structures. A more detailed summary of the peptide bond geometry for the subtilisin Novo and CI-2 molecules is given in Table 2.4. The mean values of the observed parameters all deviate from the expected values by less than 0.5 times the rms deviation. The deviation of minimum or maximum values from the mean exceeds 3 times the rms deviation in only one case, the minimum observed ω for the enzyme.

The distribution of the main-chain torsion angles ϕ and ψ in the structure of the complex can be compared with

Table 2.4

Summary of Peptide Bond Geometry

	Bond type	No.	Mean	RMS deviation	Minimum deviation	Maximum deviation	Expected[1] value
Subtilisin Novo	N-C$^{\alpha}$(Å)	275	1.470	0.008	1.446	1.495	1.474
	C$^{\alpha}$-C(Å)	275	1.528	0.008	1.508	1.546	1.524
	C=O(Å)	275	1.244	0.006	1.229	1.260	1.243
	C-N(Å)	274	1.321	0.006	1.306	1.337	1.323
	τ(°)	275	110.7	3.5	101.3	120.7	109.65
	ω(°)	273[2]	180.0	2.7	-169.1	172.5	180.00
CI-2	N-C$^{\alpha}$(Å)	64	1.470	0.007	1.460	1.484	1.474
	C$^{\alpha}$-C(Å)	64	1.528	0.008	1.514	1.547	1.524
	C=O(Å)	64	1.244	0.005	1.232	1.256	1.243
	C-N(Å)	63	1.321	0.006	1.311	1.333	1.323
	τ(°)	64	110.6	3.6	102.7	117.7	109.65
	ω(°)	63	179.6	2.8	-175.0	173.3	180.00

[1] Expected values from Arnott & Dover (1967).
[2] The peptide bond between Tyr167 and Pro168 is *cis*.

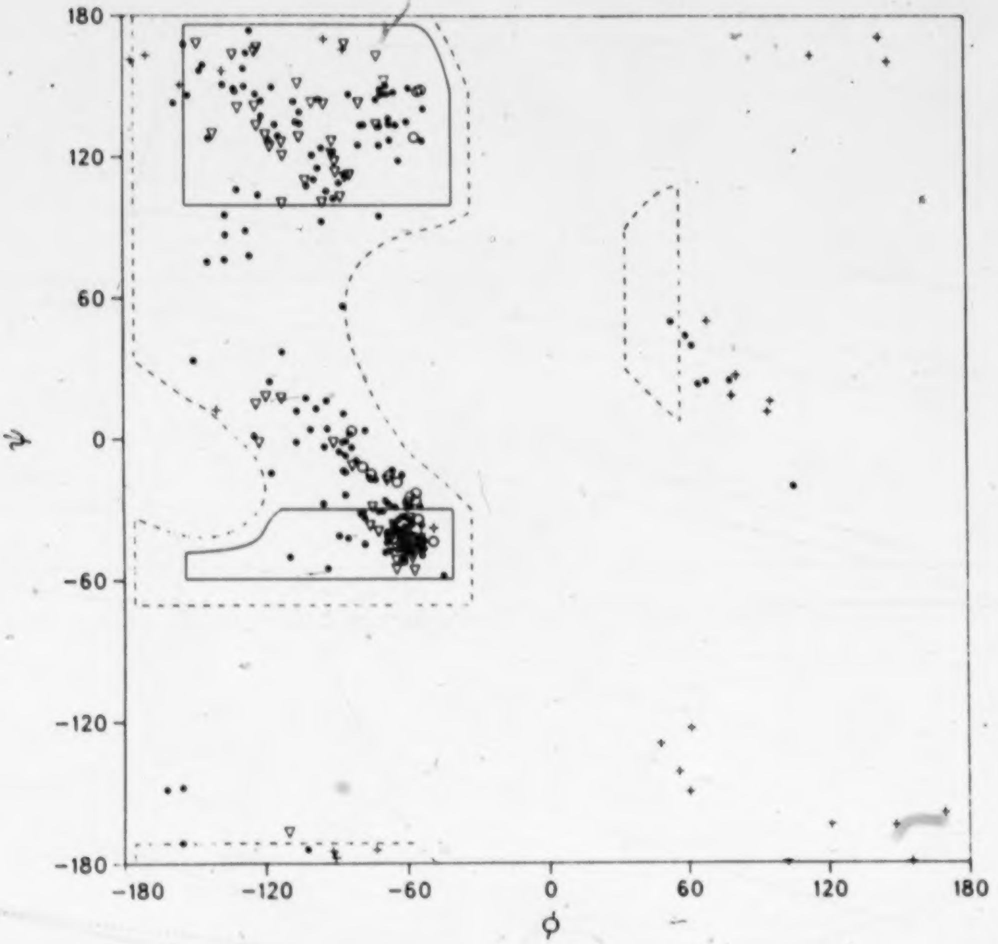


Figure 2.9. ϕ-ψ Plot for Subtilisin Novo. The solid lines enclose the fully allowed conformational regions for τ (N-C^{α}-C) of 110°; the dotted lines enclose the more permissive regions of smaller acceptable van der Waals' contacts for τ of 115° (Ramakrishnan & Ramachandran, 1965). The symbols denote prolines (o), β-branched amino acids (Δ), glycines (+), and all other amino acids (•). Residues outside the allowed conformational regions and their ϕ-ψ angles are Lys12 (77,27), Asn25 (65,21), Asp32 (-159,-148), Ser63 (106,-22), Asn77 (-153,-148), Val81 (-110,-168), Asn118 (62,38), Asn184 (69,23), and Ser204 (58,45).

allowed conformations, calculated from minimum contact radii of non-bonded atoms, in the familiar Ramachandran plot (Ramakrishnan & Ramachandran, 1965) (Fig. 2.9 and Fig.

2.10). Subtilisin Novo has 9 non-glycine residues that fall outside the allowed conformational regions; all are part of irregular structural features on the surface of the molecule such as turns or kinks. Five of these are in segments of chain with high *B*-factors and their geometry may be more poorly defined. One of them is Asp32, the catalytic aspartic acid, with ϕ = -159° and ψ = -148°, and one is Ser63, next to the catalytic histidine. In these latter 2 cases, the unusual torsion angles may be due to strain imposed by stringent structural requirements, such as the precise positioning of the catalytic side-chains. Two others are Asn77 and Val81, both ligands for Ca^{2+} ion site 1. CI-2 has 1 non-glycine residue outside the allowed regions, Asp74I; it is part of a highly mobile irregular turn at the edge of the inhibitor. The torsion angles for most residues fall well within the allowed regions for both molecules. A major cluster of conformations can be seen in the region for right-handed α-helix (-62°, -41°, Baker & Hubbard, 1984) for subtilisin Novo. The average values for residues in α-helices in subtilisin Novo are ϕ = -64° and ψ = -39°.

The final electron density map calculated for the subtilisin Novo:CI-2 complex at cycle 71 of the refinement is generally of high quality. Some regions of unsatisfactory density are still present; these are described in Table 2.5. All residues were examined for weak density at a contour level of $0.56e/Å^3$, 10% of the maximum density in the final

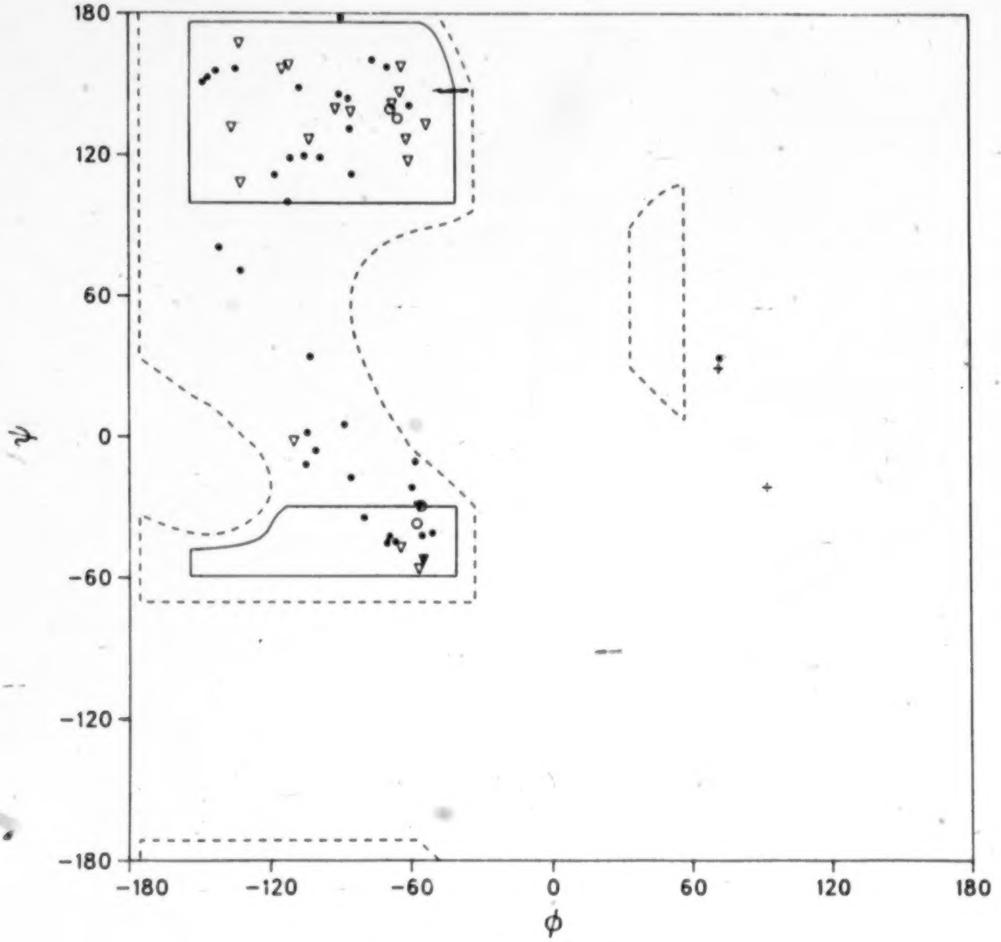


Figure 2.10. ϕ-ψ Plot for CI-2. The solid lines enclose the fully allowed conformational regions for τ (N-C$^{\alpha}$-C) of 110°; the dotted lines enclose the more permissive regions of smaller acceptable van der Waals' contacts for τ of 115°. (Ramakrishnan & Ramachandran, 1965). The symbols denote prolines (o), β-branched amino acids (Δ), glycines (+), and all other amino acids (•). The residue with torsional angles outside the allowed regions is Asp74I (71,34).

map. Not surprisingly, most residues with weak or missing density are on the surface of the enzyme or the inhibitor, where there is more freedom for motional disorder. The possible alternate side-chain conformations noted in Table

2.5 have been included in the model by placing water molecules in them. Examples of regions of good and poor density are given in Figure 2.11. Fig. 2.11a shows part of the active site of the enzyme, including the P_2, P_1, and P_1' residues of the inhibitor. Fig. 2.11b shows a segment of the only α-helix in the inhibitor; 4 of the 5 residues shown are charged and lie on the hydrophilic side of the helix with their side-chains pointing into the surrounding solvent.

One final number is generally quoted to indicate the overall quality of a crystallographic structure. This is the *R*-factor, which was defined at the beginning of this chapter. The advantage of using the *R*-factor as a measure of quality is that it is a measure familiar to crystallographers and readers of crystal structure papers; crystallographers usually have a qualitative idea of the relationship between the soundness of a structure and its *R*-factor. The *R*-factor is not, however, a valid statistical measure of the agreement between the observed and calculated structure factors. In addition, all workers do not use the same criteria for the selection of data to calculate their *R*-factors, so that the values quoted for different structures are not directly comparable. Alternate measures such as the correlation coefficient (r) have been proposed, but have not been widely calculated or reported. The *R*-factor for the structure of the complex between subtilisin Novo and CI-2, for all data in the resolution range 8.0 to 2.1Å with

Table 2.5

Residues with Poor Electron Density at Cycle 71

Residue	Comment
Ser9	Noisy density for O^{γ}, possible thermal disorder.
Lys12	No density past C^{δ}.
Asn25	Weak alternate side-chain conformation with χ_1 rotated by 180°.
Lys27	No density for C^{ϵ}, N^{ζ}.
Lys43	No density for C^{ϵ}, weak for C^{γ} & N^{ζ}.
Thr55	No density for C^{γ} & O^{γ}, weak for C^{α}.
Gln59	No density beyond C^{β}.
Ser161	Weak density for C^{α}, none for C^{β}.
Gln185	Weak density for C^{δ}, none for C^{γ} & carboxyl oxygens.
Pro201	Chain break between C^{α} & C.
Lys213	No density beyond C^{γ}.
Tyr217	Weak density along sides of phenyl ring.
Met222	No density for C^{γ}.
Asn240	Weak density for side-chain amide, possible alternate conformation with χ_1 rotated by 120°.
Lys256	No density for C^{γ}, C^{δ}, N^{ζ}.
Asp259	No density for side-chain carboxyl oxygens.
Lys265	Possible alternate conformation with χ_4 rotated by 180°.
Gln275	No density beyond C^{γ}.
Leu20I	No density except at the carbonyl oxygen.
Lys21I	Weak side-chain density.
Glu33I	No density beyond C^{β}.
Glu34I	No density for C^{γ} & C^{δ}.
Lys36I	No density for C^{ϵ} & N^{ζ}.
Lys37I	No density beyond C^{γ}.
Gln41I	No density beyond C^{β}.
Ile63I	No density beyond C^{β}.
Leu73I	No density for C^{δ}s.
Glu78I	Possible alternate conformation with χ_1 rotated by 180°.
Arg81I	No density beyond C^{γ}.

$I \geq \sigma(I)$, is 0.154. This value was calculated with 16128 reflections, 86% of the unique set of reflections measured. The R-factor calculated on the complete unique set is 0.211.

In summary, the structure of this complex is of high quality as measured by a number of criteria. The shifts in

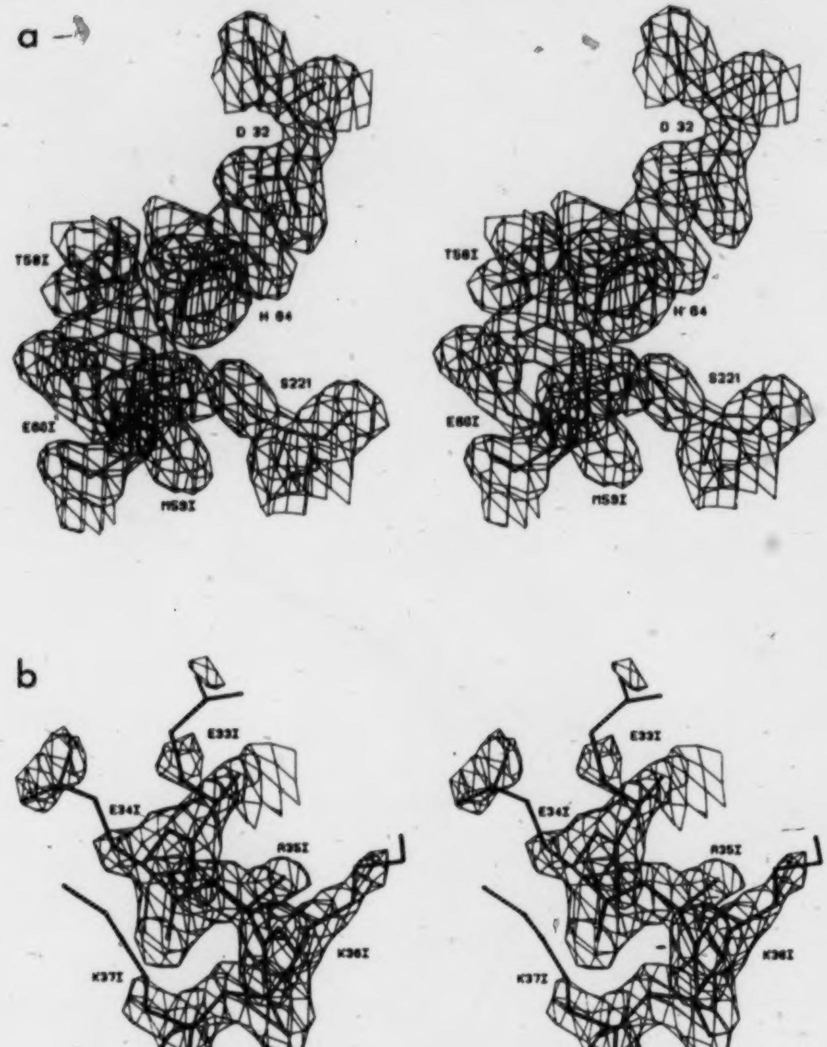


Figure 2.11. Regions of Good and Poor Electron Density. Both figures are contoured at a level of 0.56e/Å^3.
a) Residues Asp32, His64, and Ser221 from the subtilisin Novo active site, and Thr58I to Glu60I from the reactive site loop of the inhibitor. This region has some of the strongest and cleanest electron density in the map.
b) Residues Glu33I to Lys37I from the α-helix of CI-2. The side-chains of these residues protrude into bulk solvent and are highly disordered. The density for the main-chain in this region is still strong and clear.

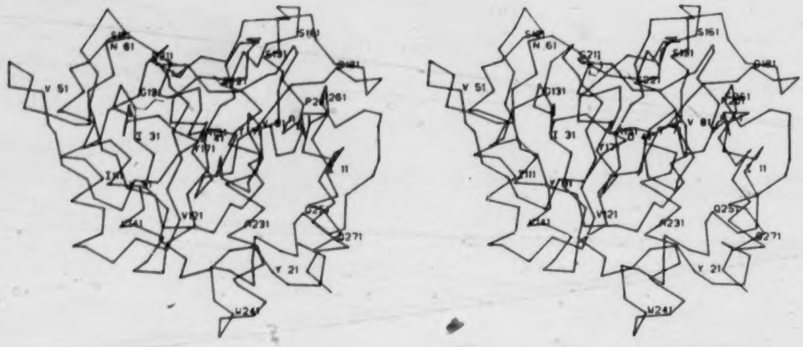
atomic parameters from the final cycle of least-squares refinement are small and a difference electron density map calculated at this stage indicates that few adjustments need to be made to the model, thus the refinement appears to have converged. The estimated errors in atomic positions are also small, and are comparable to the errors measured for other high-resolution fully-refined protein structures. The geometric parameters of the structure do not deviate significantly from ideal values; the variable main-chain torsion angles lie within or very close to theoretically allowed conformational ranges. Finally, the agreement between observed and calculated structure factors is very good for the resolution range of data that could be collected from the crystals of the complex.

2.2 The Structure of Subtilisin Novo

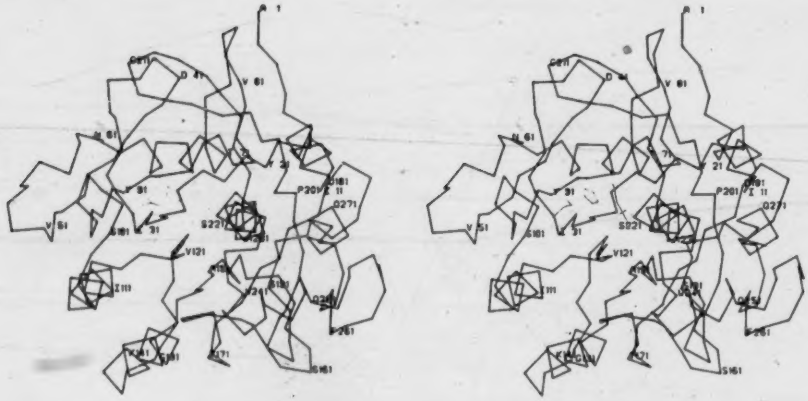
2.2.1 Secondary and Tertiary Structures

The subtilisin Novo molecule is a globular protein that is approximately heart-shaped. The active site is located in a relatively shallow groove on the surface of the molecule between the 2 lobes of the heart (Fig. 2.12a). The basic folding of the subtilisin Novo molecule in this complex with the inhibitor CI-2 is the same as that described for the unrefined structure of subtilisin Novo alone (Wright *et al.*, 1969; Drenth *et al.*, 1972). The molecule is divided into 2 segments by a central 7-stranded parallel β-sheet

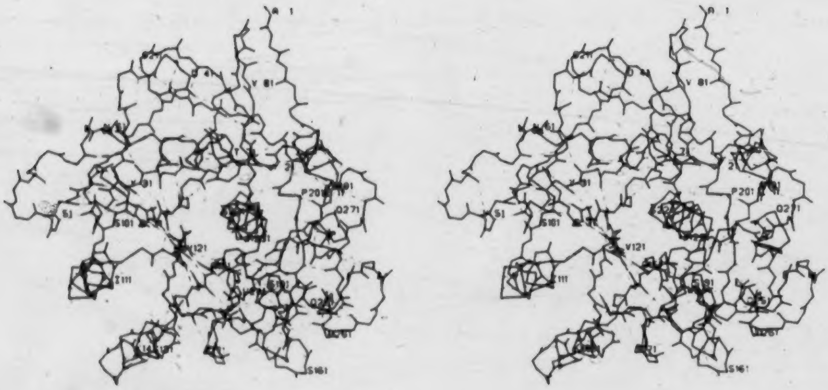
a


V 51
N 61
I 31
V121
Y 21
H241
I 11
Q271
S161

b


A 1
V 81
N 61
V 51
P201
I 11
Q271
S221
V121
I111
S161

c


A 1
V 81
P201
Q271
V121
I111
S161

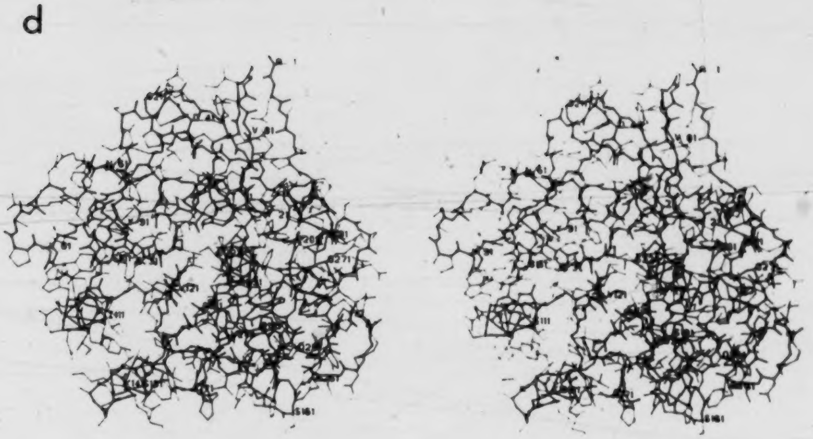


Figure 2.12. Views of Subtilisin Novo. a) An α-carbon backbone drawing of the refined subtilisin Novo molecule from the complex with CI-2. Side-chains are shown for the catalytic triad, Asp32, His64, and Ser221. The active site is in the groove at the top of the molecule. b) The molecule rotated 90° down from view (a), so that the active site is now at the front. c) The molecule in the same orientation as view (b), showing all main-chain atoms and the hydrogen bonds between them (dashed lines). d) The molecule in the same orientation as view (b), with side-chains added for all residues. Thick lines denote the main-chain, thin lines the side-chains. Every 10th α-carbon is labelled in each view.

(Fig. 2.12b, Fig. 2.12c). The larger segment contains a bundle of 7 α-helices packed against one face of the β-sheet, with 5 of the helices running approximately antiparallel to the strands of the sheet. The smaller segment contains the remaining 2 α-helices, that also pack against the β-sheet, antiparallel to it. According to the protein taxonomy of Richardson (1981), subtilisin has a doubly-wound parallel α/β structure, similar to many of the classic nucleotide binding proteins. The active site of subtilisin Novo is located similarly to the nucleotide binding sites of these other proteins, at one edge of the central sheet close to the point where the winding of the sheet changes direction. A schematic representation of the organization of secondary structural elements in subtilisin Novo is shown in Figure 2.13. The residues that make up each of these elements are listed in Table 2.6.

Secondary structural elements are generally defined by hydrogen-bonding patterns and ϕ-ψ torsional angles for the main-chain. Hydrogen atoms are not often included in the refinement of protein structures from X-ray diffraction data, so their positions must be inferred by other means. Two methods were used in the assignment of hydrogen bonds and subsequently the definition of secondary structural elements for the subtilisin Novo and CI-2 molecules. The

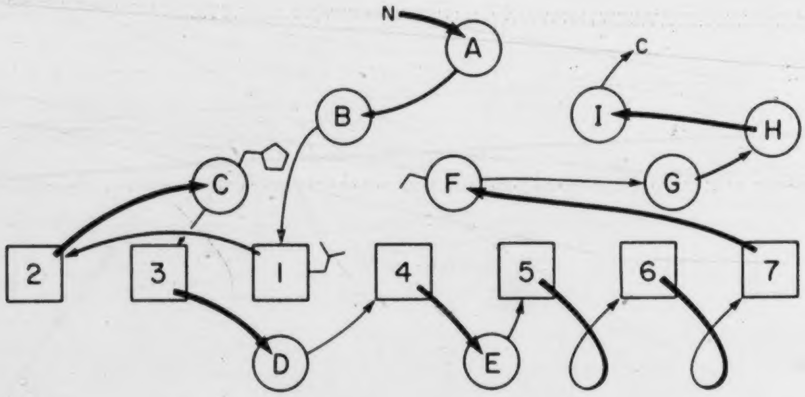


Figure 2.13. Topology of Subtilisin Novo. Strands of the central parallel β-sheet are numbered and indicated by boxes. The α-helices are indicated by circles and capital letters. The side-chains of the catalytic triad, Asp32, His64, and Ser221, show the approximate position of the active site.

first method requires the calculation of hydrogen atom positions based on standard geometry. Hydrogen bonds were defined by three parameters; distance from donor to acceptor (maximum 3.40Å), distance from proton to acceptor (maximum 2.40Å), and donor-proton-acceptor angle (minimum 125.0°). Since the hydrogen atom position is ambiguous for serine, threonine and tyrosine hydroxyl groups, and lysine and N-terminal amino groups, hydrogen bonds involving these residues were chosen solely on the basis of donor-acceptor distance. From the list of hydrogen bonds defined by this method secondary structural elements were chosen according to characteristic hydrogen-bonding patterns and ϕ-ψ angles. Visual inspection of the structure on the computer graphics system was used to check all assignments.

The second method used to define secondary structure was that of Kabsch & Sander (1983). With this method, hydrogen bonds are assigned based on the electrostatic interaction energy between donor and acceptor. Secondary structural elements are then built up from elementary hydrogen-bonding patterns. The agreement between these two methods was very good. In the few cases where the two methods did not agree, the secondary structure was assigned by the geometric criteria. Most points of disagreement occurred at the beginning or end of secondary structural elements, where the geometry of hydrogen bonds is frequently poor. Hydrogen bonds involving side-chain atoms are not considered by the Kabsch & Sander program, so these were

Table 2.6

Secondary Structural Elements of Subtilisin Novo

Element	Residues Involved
Parallel β-sheet	(1) Val26-Asp32, (2) Gly46-Met50, (3) Ser89-Val95, (4) Asp120-Met124, (5) Val148-Ala153, (6) Ile175-Val180, (7) Val198-Gly202.
Antiparallel β-sheet	Ile205-Leu209, Lys213-Tyr217.
β-bridges[1]	Gly128, Tyr167, Ala179-Asp181, Gln185-Ala187, Thr255, Leu267.
α-helices	(A) Tyr6-Ile11, (B) Ala13-Ser18, (C) Ser63-Ala74, (D) Tyr104-Asn117, (E) Ser132-Ser145, (F) Ser224-His238, (G) Thr242-Asn252, (H) Asp259-Gly264, (I) Asn269-Ala274.
Helical turns	Thr220-Ser224, Asp259-Tyr263.
Type I turns[2]	Ser9-Lys12, His17-Gly20, Asp36-His39, His39-Leu42, Val51-Glu54, Ala85-Ala88, Gly97-Gly100, Ile115-Asn118, Val143-Gly146, Tyr171-Val174, Asp181-Asn184, Ala187-Ser190, Gly193-Leu196, His238-Trp241, Ala272-Gln275.
Type II turn	Gly23-Val26.
Type II' turns	Ser159-Ser162, Tyr263-Gly266.
Type III turns	Ala13-Leu16, Leu16-Gln19, Gln103-Trp106, Pro168-Tyr171, Gly219-Met222, Ala223-His226, Ser224-Val227, Asp259-Tyr262.

[1] β structure involving only 2 hydrogen bonds (Kabsch & Sander, 1983).

[2] The turn type is assigned on the basis of the ϕ_2-ψ_2 and ϕ_3-ψ_3 angles (Crawford *et al.*, 1973).

assigned solely by the geometric criteria.

The hydrogen bonds defined by the criteria given above can be grouped into several classes according to the type of secondary structure they belong to and whether side-chain or solvent atoms participate in the bond. The mean geometric parameters for the hydrogen bonds of each of these classes are summarized in Table 2.7. The mean values of the distances and angles are the same for the α-helices, parallel

Table 2.7

Hydrogen-bonding Parameters for Subtilisin Novo

Type	No.[1]	N···O(Å) (O···O)[2]	H···O(Å)	N-Ĥ-O(°)
Main-chain to main-chain				
α-Helix(5→1)	51	2.99(0.14)	2.06(0.15)	154(9)
Antiparallel β-sheet	12	2.92(0.10)	1.98(0.12)	157(11)
Parallel β-sheet	30	2.93(0.15)	1.97(0.16)	162(8)
3_{10} turns(4→1)	26	3.05(0.13)	2.18(0.11)	147(12)
Main-chain to side-chain				
N-H···O	32	2.96(0.16)	2.04(0.19)	155(12)
C=O···H-N	10	2.89(0.10)	1.97(0.12)	155(14)
C=O···H-O	26	2.86(0.29)	-[3]	-
Side-chain to side-chain				
N-H···O	7	2.93(0.22)	2.00(0.21)	155(9)
O(H)···(H)O	14	2.87(0.25)	-	-
Main-chain to solvent				
N-H···O	31	2.99(0.20)	2.06(0.21)	156(12)
C=O···O	73	2.92(0.34)	-	-
Side-chain to solvent				
N-H···O	11	2.92(0.26)	2.04(0.26)	147(11)
O(H)···(H)O	35	2.88(0.26)	-	-

[1] Limits for acceptance of a hydrogen bond: donor-acceptor distance ≤ 3.40Å, hydrogen-acceptor distance ≤ 2.40Å, donor-hydrogen-acceptor angle ≥ 125°.

[2] Values in parentheses are the sample standard deviations from the mean.

[3] Values are not given for donors with ambiguous hydrogen atom positions.

and antiparallel β-sheets, within the limits of accuracy of the structure. The two types of turns listed have somewhat lower average values for the donor-hydrogen-acceptor angle, but the average depends strongly on the lowest angle accepted as being a hydrogen bond. The lower limit of 125°

used here is a loose criterion. The variance of this angle for the turns is also somewhat larger than for the more regular and extended types of secondary structure. The mean values of the distances given are statistically the same for intra-main-chain hydrogen bonds and those involving side-chains or solvent molecules. Results similar to these have been reported for other fully-refined protein structures: penicillopepsin (James & Sielecki, 1983~~),~~ actinidin (Baker, 1980), and human lysozyme (Artymiuk & Blake, 1981).

The following paragraphs contain a more detailed description of the tertiary structure of the subtilisin Novo molecule, in terms of the characteristics of and relationships between the secondary structural elements of Table 2.6. The discussion may be understood more easily with frequent reference to Figures 2.12 and 2.13. The N-terminus of subtilisin Novo lies on the surface of the molecule and begins a short stretch of extended chain that runs alongside a large surface loop. The loop (residues Leu75 to Gly83) is the binding site for the first of two Ca^{2+} ions associated with the enzyme and the side-chain of Gln2 is one of the ligands to this Ca^{2+}. Helix A follows the extended chain; it is 1.5 turns long and also lies on the molecule's surface. All the surface helices in subtilisin Novo have some amphipathic character, with a propensity for polar residues on their outer face and non-polar residues on the face toward the protein interior. A type I turn leads into helix B, which lies at an 82° angle to helix

A; B is also short and amphipathic.

An extended loop (residues Gln19 to Asn25) traverses the enzyme surface and makes a 180° turn to reach the start of strand 1 of the central parallel β-sheet. All but one of the strands of the parallel β-sheet are buried in the interior of the protein; strand 2 is somewhat exposed on the surface, as it forms one edge of the sheet. The other edge of the sheet is buried by segments of extended chain. Parallel β-sheets in general are well-protected from solvent by other segments of main-chain covering both sides of the sheet. On the basis of observations of similarly well-protected parallel β-sheets in a number of proteins, Richardson (1981) has proposed that parallel β-sheet is inherently less stable than antiparallel sheet and requires more protection for its hydrogen-bonding network.

Strand 1 of the central sheet runs up through the molecule toward the active site, antiparallel to helix B but 20 angstroms away from it. Asp32, the essential aspartic acid of the catalytic triad in the subtilisins, is located at the C-terminus of strand 1. A large irregular surface loop (residues Ser33 to Ala45) brings the chain around 360° and back to the N-terminal edge of the sheet for the start of strand 2. This loop lies against the loop binding Ca^{2+} ion 1, on the opposite side from the N-terminus of the enzyme. It contributes the second ligand for the Ca^{2+} ion, the side-chain of Asp41. Strand 2 of the sheet is only 5 residues long (Gly46 to Met50) and is partly exposed to bulk solvent. Another large irregular surface loop follows strand 2 (Val51 to Asn62); this loop contains Thr55, which is the single deletion in subtilisin Carlsberg relative to subtilisin Novo (determined on the basis of three-dimensional structural homology). There is a very unusual left-handed crossover connection between strands 2 and 3 in subtilisin Novo (Richardson, 1976).

Helix C begins in the active site, with His64 of the catalytic triad located at its N-terminus. This helix is predominantly buried and lies antiparallel to strands 1 and 2 of the sheet. The C-terminal residue of helix C is Ala74 but, based on hydrogen-bonding parameters and ϕ-ψ angles, residues Val84 and Ala85 can be considered as a continuation of the helix. An interrupted helix such as this is very rare in known protein structures (Richardson, 1981). The segment of chain that interrupts the helix is the loop Leu75 to Gly83, the Ca^{2+}-binding loop discussed above. Secreted serine proteinases are characteristically stabilized by Ca^{2+}, often to provide protection against proteolytic degradation or autolysis (Kretsinger, 1976). Kretsinger has proposed that this calcium requirement prevents these proteinases from being active inside the cell, where Ca^{2+} concentrations are usually very low. Subtilisin Novo is stabilized against autolysis by Ca^{2+} (Matsubara *et al.*, 1958). A role for the Ca^{2+}-binding loop observed here in the stabilization of the subtilisin Novo molecule can be proposed readily. Removal of the Ca^{2+} ion from the loop could disrupt its structure and at the same time disrupt the structure of helix C, to which it is intimately attached. This helix runs through the hydrophobic core of the enzyme and contains one of the essential active site residues (His64). A disturbance of the hydrophobic core could alter the folding of the protein and leave it vulnerable to proteolytic degradation. A disturbance of the geometry of the active site residues would likely render the enzyme inactive; this may be the situation of the enzyme before it is secreted. The known instability of subtilisin Novo below pH 5 (Ottesen & Spector, 1960) could be due to protonation of Asp41 and disruption of this Ca^{2+} binding site.

Helix C is followed by a 180° turn to strand 3 of the parallel β-sheet. This strand runs back up toward the active site, antiparallel to helix C. A reverse turn at the end of the strand leads into a segment of extended chain, residues Leu96 to Gln103. Residues in the first half of this segment form part of the side-chain binding pocket for the P_2 residue of a substrate or inhibitor; the second half forms a short stretch of antiparallel β-sheet with the P_2 to P_4 residues before turning to the start of helix D. Helix D is the first of the two helices that pack on the opposite face of the central sheet to the other six helices. It lies on the surface of the molecule, antiparallel to strands 1 and 3, and at an angle of 31° to helix C. Although all the helices packed against the parallel sheet run antiparallel to it, they are not all parallel to each other due to the characteristic twist of the sheet. The sheet in subtilisin Novo is strongly twisted, such that the strands at one edge of the sheet are perpendicular to the strands at the other edge.

Looking down the axes of helices D and F (Fig. 2.12), a plane can be drawn through these two helices that divides the molecule roughly in half. All the residues from the N-terminus through helix D pack together to make up one half of the molecule, and most of the remaining residues make up the second half of the molecule. Since subtilisin Novo is a globular protein, however, there is extensive contact at the interface of the two halves and they cannot be considered to be separate 'domains' in the usual sense of the term.

A 180° turn and strand 4 follow helix D. Strand 4 is the point of reversal of the β-sheet winding, and lies beside strand 1 in the core of the molecule. This strand ends in another extended chain region that interacts with substrates or inhibitors in the active site; Gly127 forms two antiparallel β-sheet bonds with the P_3 residue. The main-chain of residues Ser125 to Gly128 forms one side of the P_1 specificity pocket on the enzyme surface. Another turn begins helix E, which packs on the same side of the central sheet and parallel to helix D. A final 180° turn leads to strand 5, which again runs back toward the active site. The C-terminal residues of this strand form the other wall of the P_1 specificity pocket and contribute to the oxyanion hole, the polarizing pocket for the carbonyl oxygen of the P_1 residue.

A large extended surface loop, containing several kinks, eventually turns almost 360° to the start of strand 6. This loop protects strands 5 and 6 from contact with solvent, and portions of it also lie against the side of helix E. Because it lies on the surface and is very irregular in structure, portions of the loop are disordered and have poor electron density (Fig. 2.11a and Table 2.5). Residues Gly169, Tyr171, and Val174 from this loop are three of the ligands binding the second Ca^{2+} ion in subtilisin Novo. A function for Ca^{2+} ion 2 in stabilizing the protein is more difficult to visualize than for ion 1, since the second binding site is not directly connected to the active site and the protein chain around it is already irregular. The single *cis*-peptide bond in subtilisin Novo is contained in the loop, between Tyr167 and Pro168; the *cis* bond is between the i+1 and i+2 residues of a 5-membered turn on the surface of the enzyme.

Strand 6 of the parallel β-sheet is followed by another large irregular loop that packs against the loop between strands 5 and 6. Only one edge of this loop is on the protein surface. Strand 7, the final strand of the sheet, is followed by a long twisted loop of antiparallel β-sheet that crosses over from the second half of the molecule to the first half and then back again. This 'arm' is in contact with helix C and the loop binding Ca^{2+} ion 1. Helix F comes after the antiparallel loop and runs down the center of the molecule across its entire diameter. It is the central helix of the 6-helix cluster on one side of the parallel β-sheet, and is at angles of 10° and 23° to helices D and E on the other side of the sheet. Helix C is at an angle of 35° to F. The N-terminus of helix F contains the remaining residues forming the oxyanion hole, as well as the third member of the catalytic triad, Ser221.

A short surface loop after helix F (residues Pro239 to W241) reverses the direction of the polypeptide chain and is followed by helix G. The helix axis of G is at an angle of 64° to that of F but, due to the twist of the β-sheet, G is still antiparallel to strand 5 of the sheet and in contact with strands 6 and 7. Residues Asn252 to Asn269 forms the final extended surface loop that stacks against the loop between strands 6 and 7 and covers the edge of the parallel β-sheet. The interactions between this loop and strand 7 are too irregular to be considered an extension of the sheet. Helix H is in the middle of this loop; it is composed of residues Asp259 to Gly264 and is 1.5 turns long. It lies on the surface of the molecule. Helix I terminates the molecule; it is short and amphipathic. The C-terminus of subtilisin Novo is thus on the surface of the molecule, at a distance of about 34Å from the N-terminus. Both termini are distant from the active site.

The observed distribution of angles between the axes of helices packing together in proteins has been explained by a 'ridges into-grooves' model (Chothia *et al.*, 1981), with the ridges formed by residues on the surface of one helix packing into the grooves between the ridges on an adjacent helix. The angle between two helix axes is calculated by projecting the two helices onto their plane of contact. Positive angles correspond to counterclockwise rotations of the top helix relative to the bottom one. The distribution of observed packing angles has a maximum at -50° and another at +20°, although the distribution is spread over a wide range of possible values. The observed inter-helix angles in the subtilisin Novo structure fit the general distribution fairly well. Within the bundle of 7 helices on one side of the β-sheet, the angles range from -31° (F and I) to -64° (F and G). The angle between B and F is +35°. The mean value for the angles clustered around -50° is -46°. The angle between D and E, on the other side of the β-sheet, is +14°.

2.2.2 Calcium Binding Sites

Among the first set of solvent molecules chosen from a difference map during refinement of the subtilisin Novo:CI-2 complex were two whose occupancies stayed at 1.0 and whose B-factors dropped rapidly to 2, the lowest value allowed by the local version of the refinement program. Since the crystallization medium for this complex contains KH_2PO_4 and subtilisin Novo is known to be stabilized by divalent cations, these two solvent molecules were suspected to be either K^+ or Ca^{2+} ions. In subsequent refinement cycles

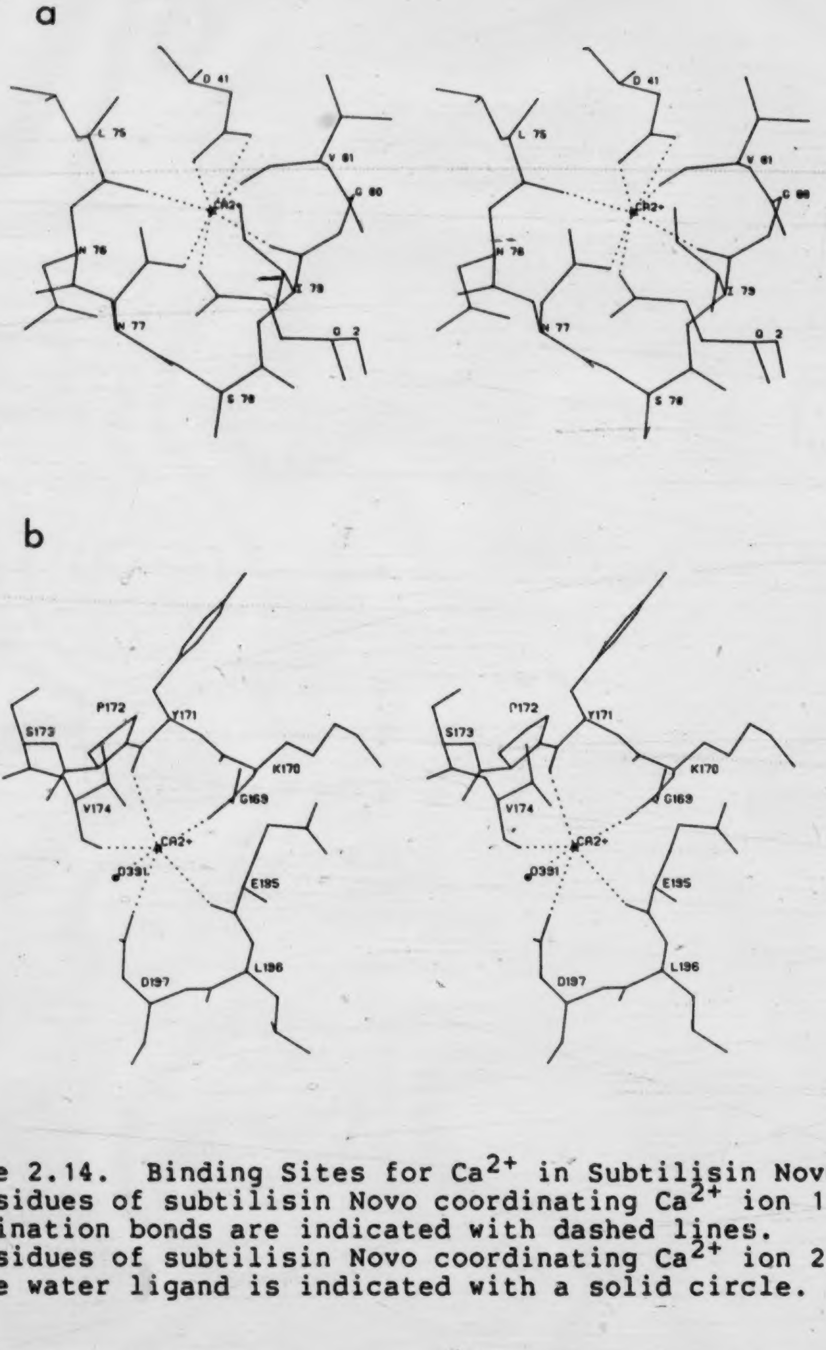

a
D 41
L 75
V 81
G 80
CA2+
N 76
I 79
N 77
O 2
S 78
b
P172
Y171
S173
K170
G169
V174
CA2+
O391
E195
L196
D197

Figure 2.14. Binding Sites for Ca^{2+} in Subtilisin Novo. a) Residues of subtilisin Novo coordinating Ca^{2+} ion 1. Coordination bonds are indicated with dashed lines. b) Residues of subtilisin Novo coordinating Ca^{2+} ion 2. The single water ligand is indicated with a solid circle.

these two molecules were treated as sulphur atoms, with no restraints on non-bonded contact distances. Sulphur was considered a reasonable temporary model for a large ion of uncertain chemical nature. Near the end of the refinement, examination of difference maps, the coordination geometry of the ions, the nature of their ligands, and the ion-ligand distances indicated that the ions were probably Ca^{2+} and they were refined as such for the last 7 cycles. The final *B*-factor and occupancy for ion 1 are 10Å² and 0.97, and for ion 2 are 28Å² and 0.72, indicating that ion 1 is better defined and probably bound more tightly to the enzyme. Although no Ca^{2+} is added to the crystallization medium, the enzyme may be purified with the ions attached, especially if they are tightly bound and necessary for stablity. In the structure of the native subtilisin Novo (Drenth *et al.*, 1972), high peaks of electron density were also found in positions close to these two ions. The two Ca^{2+} ions and their ligands are shown in Figure 2.14, and their coordination geometry is described in Table 2.8.

The average distance of the coordinating ligands to Ca^{2+} ion 1 is 2.40Å, and the average for Ca^{2+} ion 2 is 2.91Å. A typical Ca^{2+}-O distance in small molecule crystal structures is 2.44Å (Brown & Shannon, 1973), and coordination by 6 or 7 ligands in a roughly octahedral arrangement is common for Ca^{2+} ions bound to carboxyl and carbonyl ligands (Einspahr & Bugg, 1981, 1984). The coordination bond lengths for Ca^{2+} ion 2 are longer than the typical distances; this may indicate that this ion is actually K^{+}, for which the typical coordination bond length

Table 2.8

Geometry of Calcium Ion Binding Sites

Coordinating atoms	Ca^{2+} - O bond length (Å)	Ca^{2+} - $\hat{O}$ - C bond angle (°)
Ca^{2+} #1		
Gln2 $O^{\epsilon 1}$	2.38	143
Asp41 $O^{\delta 1}$	2.43	96
Asp41 $O^{\delta 2}$	2.56	90
Leu75 O	2.35	148
Asn77 $O^{\delta 1}$	2.43	123
Ile79 O	2.35	156
Val81 O	2.31	162
Ca^{2+} #2		
Gly169 O	2.85	167
Tyr171 O	3.01	116
Val174 O	2.83	153
Glu195 O	3.04	144
Asp197 $O^{\delta 2}$	2.81	136
O391 O	2.92	-

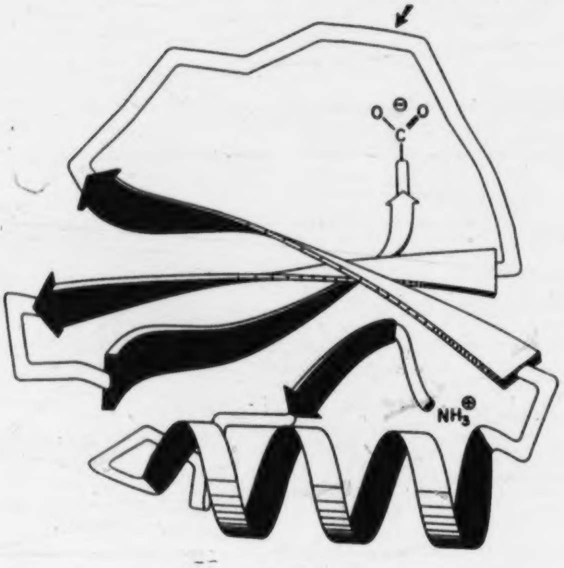
is 2.82Å (Brown & Shannon, 1973). Whether ion 2 is Ca^{2+} or K^+, its final atomic parameters would remain the same because the two types of ions are isoelectronic and have very similar scattering factors in the relevant resolution range. Both ion binding sites in subtilisin Novo are unusual in one respect compared to other Ca^{2+} binding sites in proteins. Einspahr & Bugg (1984) found that most of the ligands in Ca^{2+} binding sites come from a single segment of less than 13 residues in the protein sequence; the ligands in subtilisin Novo come from widely separated segments of the protein sequence.

2.3 The Structure of CI-2

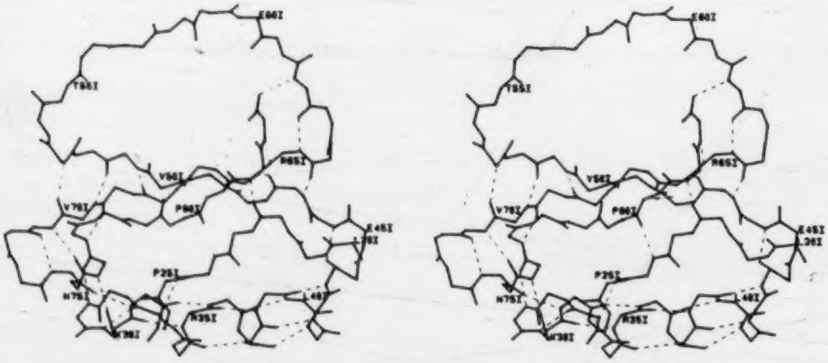
The CI-2 molecule is a wedge-shaped disk of approximate dimensions 28x27x19Å, similar in size and shape to the avian ovomucoid inhibitors (Fujinaga *et al*., 1982). The loop containing the reactive site bond (the peptide bond attacked by the catalytic apparatus of the enzyme) forms the narrow end of the wedge. The polypeptide chain folding of CI-2 is quite different from the tertiary organization of the ovomucoid inhibitors and other inhibitor families of known three-dimensional structure (Marquart *et al*., 1983; Hirono *et al*., 1984). The core of CI-2 is composed of an α-helix of 13 residues, 3.6 turns, and a 'pseudo' β-sheet of mixed parallel and antiparallel strands (see below). Figure 2.15 gives 3 representations of the folding of CI-2. Table 2.9 contains a list of the secondary structural elements of CI-2 and the hydrogen bonds made by the residues comprising those elements (Fig. 2.15b).

The first 19 residues of CI-2 (Fig. 2.2) are disordered and are not seen in the electron density map of the complex. The inhibitor is subject to hydrolysis during purification and the amino acid sequence of the inhibitor from the complex crystals indicates that the inhibitor has a 'ragged' N-terminus in the crystals (M. Carpenter & L. Smillie, unpublished results). The first residue observed in the crystal structure is Leu20I. Leu20I to Lys21I are in an extended conformation and interact little with the body of the inhibitor. Thr22I and Trp24I form a β-bridge of two

a


O
C
O
NH$_3^{\oplus}$

b


E60I
Y55I
R65I
V50I
V70I
P80I
E45I
L20I
P25I
N75I
L40I
N35I
W30I

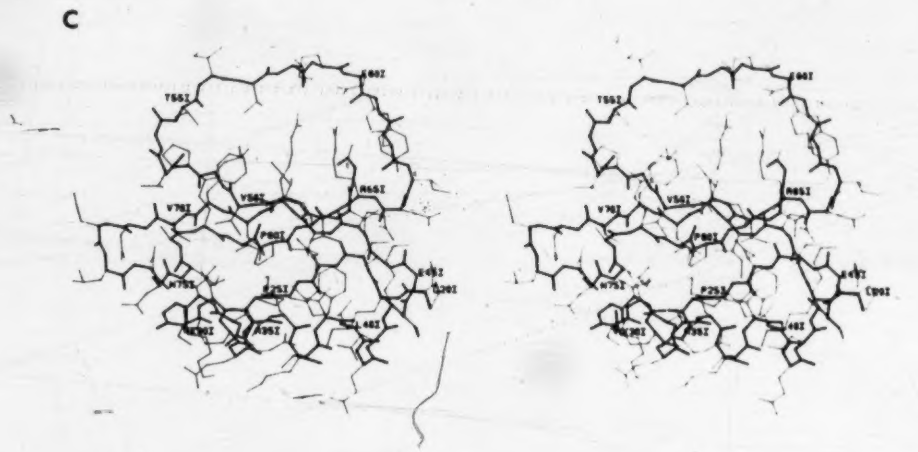


Figure 2.15. Views of CI-2. All views include only residues Leu20I to Gly83I; residues 1 to 19 are not seen in the electron density map. a) A stylized drawing of the secondary structural elements in CI-2. The small arrow at the top indicates the reactive site bond. Large arrows denote strands of β-sheet, wide ribbons are α-helix, and narrow ribbons are turns or unclassified structure. b) The main-chain atoms of CI-2 and their hydrogen-bonding interactions. Hydrogen bonds are drawn with dashed lines. c) The complete CI-2 molecule. The main-chain is drawn with heavy lines and the side-chains with light lines. Hydrogen bonds are indicated with dashed lines. Every fifth amino acid residue is labelled in views (b) and (c).

hydrogen bonds to Pro80I and Val82I near the C-terminus. They are followed by two linked 3_{10} turns that reverse the direction of the polypeptide chain and lead into the single α-helix. The i-1 and i+1 residues of the α-helix (Lys30I and Val32I) also participate in an antiparallel β-bridge to Asp74I and Ile76I near the C-terminus. The α-helix lies in the curve created by the characteristic twist of the 'pseudo' β-sheet, with its helix axis antiparallel to the direction of the two parallel strands of sheet.

Table 2.9

Secondary Structural Elements of CI-2

Element	Residues[1]	N-H···O length (Å)	N-Ĥ-O angle (°)
Parallel β-sheet	Gln47I-Asp64I	2.67	160
	Val66I-Gln47I	2.97	166
	Ile49I-Val66I	3.09	163
	Leu68I-Ile49I	2.76	170
	Leu51I-Leu68I	2.81	172
	Val70I-Leu51I	2.94	158
	Val53I-Val70I	2.86	154
Antiparallel β-bridges	Val82I-Thr22I	2.97	161
	Trp24I-Pro80I	2.96	154
	Ile76I-Lys30I	2.75	160
	Val32I-Asp74I	3.00	135
	Ala77I-Phe69I	2.70	173
	Asp71I-Asn75I	3.09	162
α-helix	Ala35I-Ser31I	2.93	163
	Lys36I-Val32I	2.96	168
	Val38I-Glu34I	3.32	161
	Ile39I-Ala35I	3.02	173
	Leu40I-Lys36I	3.02	167
	Gln41I-Lys37I	3.20	136
	Asp42I-Val38I	2.96	145
	Lys43I-Ile39I	2.59	156
Type I turns	Ala46I-Lys43I	2.98	158
	Arg65I-Arg62I	3.28	159
	Asp74I-Asp71I	2.95	165
Type II turn	Lys30I-Leu27I	2.84	151
Type III turn	Leu27I-Trp24I	3.15	162
Unclassified	Gly29I-Ile76I	2.85	166
	Arg62I-Gly83I	2.93	178
	Gly83I-Arg65I	2.98	164

[1] The residue listed first in each pair is the hydrogen bond donor.

A reverse turn follows the α-helix and starts the first strand of parallel β-sheet, residues Gln47I to Val53I. A reversal of chain direction ends this strand and leads into a wide loop of extended chain that contains the reactive site bond. The reactive site loop lies on the opposite side of the β-sheet from the α-helix and forms a large semi-circle above the sheet, with no interactions between the main-chain atoms of the loop and the main-chain atoms of the sheet. The loop is concluded by a reverse turn and the second strand of the parallel β-sheet. This strand contains Arg65I and Arg67I, residues with extended side-chains that reach across the gap between the sheet and the loop to provide part of the hydrogen-bonding network that supports the extended conformation of the loop (Fig. 2.16).

The final reverse turn begins the C-terminal strand of the β-structure. Residues Asn75I to Ala77I form an antiparallel β-bridge of two hydrogen bonds to Phe79I to Asp80I

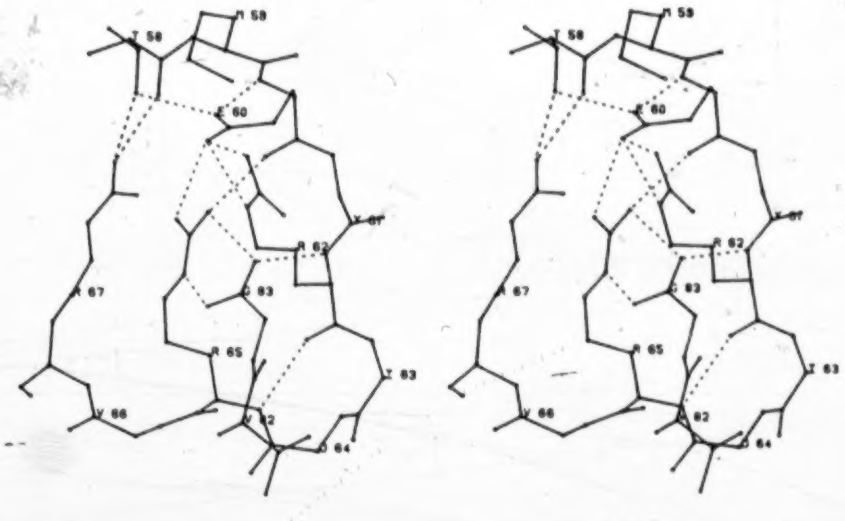


Figure 2.16. Hydrogen Bonding of the Reactive Site Loop in CI-2. Dashed lines indicate hydrogen bonds. Side-chains of residues not involved in hydrogen bonding have been omitted for clarity.

in strand 2 of the parallel sheet. From Ala77I to Gly83I, the C-terminal strand diverges from strand 2 because the main-chain is kinked after Ala77I. The remaining residues have the appropriate ϕ-ψ angles for β-sheet, but interact with strand 2 through a group of well-ordered solvent molecules lying between the two strands and participating in hydrogen-bond bridges (Fig. 2.17). The 'pseudo' β-sheet is thus composed of Thr22I to Trp24I at one edge with two antiparallel bonds to the C-terminal strand, the C-terminal strand with two antiparallel bonds to strand 2, and strands 1 and 2 of the parallel β-sheet. For convenience, these 4 strands will be referred to as a β-sheet in all following discussions, but it should be remembered that some of their hydrogen-bonding patterns are irregular.

The interface between the α-helix and the β-sheet provides the hydrophobic core of CI-2, which is composed of residues Trp24I, Leu27I, Val32I, Ala35I, Val38I, Ile39I, Leu40I, Ala46I, Ile48I, Val50I, Val66I, Leu68I, Val70I, Ile76I, Pro80I, and Val82I. The aliphatic portions of 3 lysine side-chains are also part of the core: Lys21I, Lys36I, and Lys43I. Being a small protein, CI-2 has a similar number of hydrophobic residues partially or fully exposed on its surface. The only buried polar group in the molecule is the carboxylate of Gly83I, the C-terminal residue (Fig. 2.16). The reactive site loop of this inhibitor is supported by an extensive network of hydrogen bonds and electrostatic interactions, involving the P_2 and P_1'

residues (Thr58I and Glu60I), Arg62I, Arg65I, Arg67I, and the C-terminal carboxyl group. This network is shown in Figure 2.16.

The mean geometric parameters for various classes of hydrogen bonds in CI-2 are given in Table 2.10. None of these values are statistically different from the values found in the larger subtilisin Novo molecule.

2.4 The Subtilisin Novo:CI-2 Complex

2.4.1 Inhibitor Binding Interactions

The most striking feature of the complex of subtilisin Novo and CI-2 is the apparent lack of contact between the enzyme and the inhibitor except at the reactive site loop

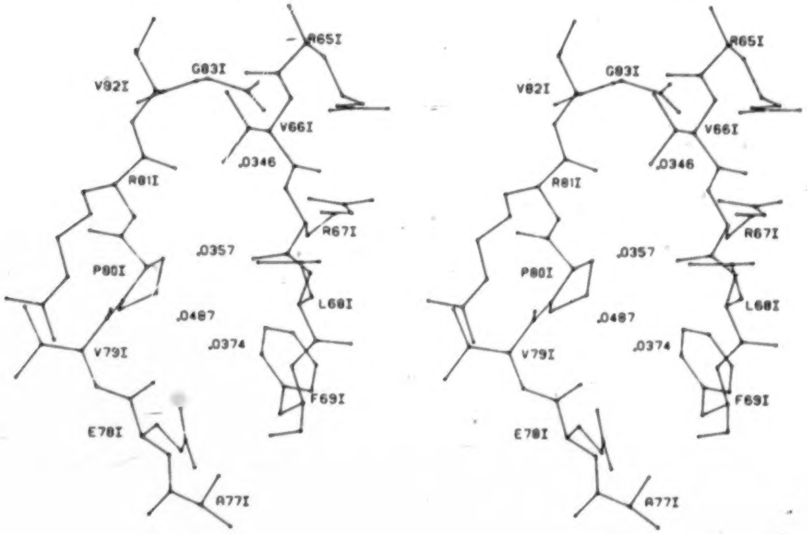


Figure 2.17. Bridging Water Molecules in the β-Sheet of CI-2. Strand 2 of the parallel β-sheet and the C-terminal strand of CI-2, with the water molecules providing hydrogen-bonding bridges between them.

Table 2.10

Hydrogen-bonding Parameters for CI-2

Type	No.[1]	N···O(Å) (O···O)[2]	H···O(Å)	N-H-O(°)
Main-chain to main-chain				
α-Helix(5→1)	8	3.00(0.21)	2.06(0.24)	159(12)
Antiparallel β-sheet	6	2.91(0.16)	1.98(0.20)	158(12)
Parallel β-sheet	7	2.87(0.14)	1.90(0.15)	163(6)
3_{10} turns(4→1)	5	3.04(0.18)	2.09(0.17)	159(5)
Main-chain to side-chain				
N-H···O	8	2.95(0.22)	2.05(0.16)	151(14)
C=O···H-N	1	2.84(0.00)	1.88(0.00)	161(0)
C=O···H-O	2	2.94(0.29)	-[3]	-
Side-chain to side-chain				
N-H···O	9	2.97(0.22)	2.05(0.25)	155(15)
O(H)···(H)O	2	2.98(0.23)	-	-
Main-chain to solvent				
N-H···O	5	3.02(0.04)	2.03(0.02)	170(8)
C=O···O	19	2.85(0.24)	-	-
Side-chain to solvent				
O(H)···(H)O	10	2.84(0.32)	-	-

[1] Limits for acceptance of a hydrogen bond: donor-acceptor ≤ 3.40Å, hydrogen-acceptor ≤ 2.40Å, donor-hydrogen-acceptor angle ≥ 125°.

[2] Values in parentheses are the sample standard deviations from the mean.

[3] Values are not given for donors with ambiguous hydrogen atom positions.

(Fig. 2.18). This phenomenon is common to most serine proteinase inhibitors of known structure (Read & James, 1986). This mode of binding allows the extended loop of the inhibitor reactive site region to interact extensively with residues in the active site cleft, but avoids possibly unfavorable interactions between residues to either side of

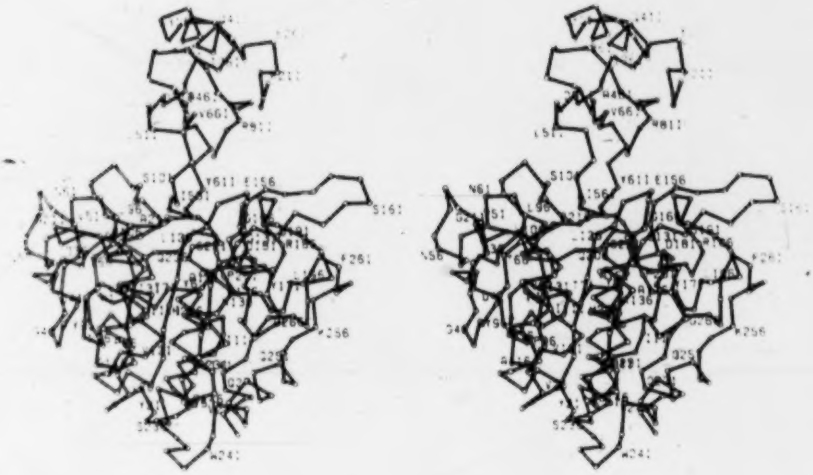


Figure 2.18. The Subtilisin Novo:CI-2 Complex. An α-carbon representation of the CI-2 molecule (filled bonds) in complex with subtilisin Novo (open bonds). Every fifth amino acid residue is labelled with the residue type and sequence number.

the active site and residues in the main body of the inhibitor. Eleven residues of CI-2 make a total of 134 contacts less than 4.0Å with 22 residues of subtilisin Novo. These contacts are summarized in Table 11.

The majority of the contacts between subtilisin Novo and CI-2 are between side-chains of residues in the reactive site loop and the enzyme residues comprising the side-chain binding pockets or recognition sites. There are far more intermolecular contacts of CI-2 with subtilisin Novo involving residues P_4 to P_1 than residues P_1' to P_3'. This distribution of contacts has also been observed in other complexes of inhibitors with serine proteinases (Huber & Bode, 1978; Read *et al.*, 1983). The active site region of

Table 2.11

Intermolecular Contacts ≤4.0Å for Subtilisin Novo and CI-2

	Ile 49I P_{11}	Leu 51I P_9	Gly 54I P_6	Thr 55I P_5	Ile 56I P_4	Val 57I P_3	Thr 58I P_2	Met 59I P_1	Glu 60I P_1'	Tyr 61I P_2'	Arg 62I P_3'	Σ
His64							10	1	3			14
Leu96					1		1					2
Asp99	1	4										5
Gly100					1	4	3					8
Ser101		2			3							5
Gly102				4	5							9
Gln103				1								1
Tyr104			3	1	5							9
Ile107					2							2
Ser125							1	1				2
Leu126					2	3		4				9
Gly127					3	5		1				9
Gly128					1							1
Ala152								2				2
Gly154								3				3
Asn155								11	3		1	15
Phe189										5		5
Tyr217											4	4
Asn218									3	5		8
Gly219								2		1		3
Thr220								4				4
Ser221								12	2			14
Σ	1	6	3	6	23	12	15	41	11	12	4	134

subtilisin Novo and the reactive site loop of CI-2 are shown together in Figure 2.19, which includes most of the residues and contacts given in Table 2.11.

The large contribution of the P_1 and P_4 residues of the inhibitor to tight binding to the enzyme is indicated by the number of their contacts to the enzyme; these two residues account for almost half the enzyme:inhibitor contacts less than 4.0Å (Table 2.11). The side-chain of the P_4 residue (Ile56I) fits neatly into the complementary S_4 hydrophobic

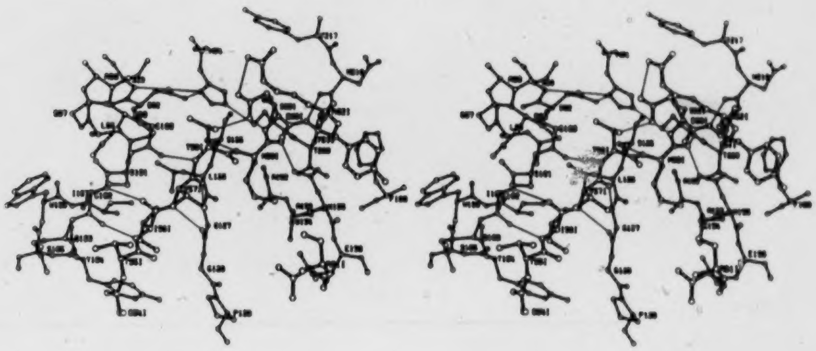


Figure 2.19. Interactions Between Subtilisin Novo and CI-2. An ORTEP drawing (Johnson, 1965) of the active site region of subtilisin Novo (open bonds) and the reactive site loop of CI-2 (filled bonds). Hydrogen bonds are indicated with thin lines. The hydrogen bond shown from the NH of Met59I (P_1) to the carbonyl-oxygen of Ser125 is long (3.38Å) and probably weak. This bond is characteristic of an incipient tetrahedral intermediate (Robertus *et al.*, 1972a). Part of the stabilization energy for the formation of a tetrahedral intermediate would be obtained by shortening, and thus strengthening, this hydrogen bond.

pocket in the enzyme, formed by the side-chains of Leu96, Tyr104, Ile107, and Leu126, and the main-chain atoms of Gly127 and Gly128. The primary binding pocket in subtilisin Novo, S_1, is similar to the analogous pocket in α-chymotrypsin in its position in the enzyme and its generally aliphatic or hydrophobic nature (Robertus *et al.*, 1972a). The pockets of both enzymes are designed to bind large or medium-sized hydrophobic residues and the nature of this S_1 pocket is the major determinant of the enzyme specificity (Kraut, 1977; Robertus *et al.*, 1972b; Markland & Smith, 1971). The atoms of the main-chain segments Ser125 to Gly127 and Ala152 to Asn155 form the walls of this pocket in subtilisin Novo.

The S_2 binding pocket in subtilisin Novo is also hydrophobic, being formed by the plane of the imidazole ring of His64 and the side-chain of Leu96. The polar O^γ of the P_2 residue, Thr58I, interacts with the hydrogen-bonding network of the inhibitor reactive site loop. The side-chain of the P_3 residue in CI-2, Val57I, points back toward the body of the inhibitor and does not interact with a specific binding site on the enzyme. This is common to all inhibitors of the serine proteinases. On the C-terminal side of the scissile bond, Glu60I (P_1') also points back toward the body of the inhibitor. This residue is a central part of the hydrogen-bonding network supporting the reactive site region (Fig. 2.16). The side-chain of the P_2' residue, Tyr61I, stacks with the side-chain of Phe189 in a hydrophobic interaction on the surface of the enzyme, rather than in a defined binding pocket. The aliphatic region of the P_3' residue side-chain, Arg62I, takes part in another hydrophobic interaction on the enzyme surface, lying close to the side-chain of Tyr217. In the subtilisin Novo:CI-2 complex, the secondary contact region defined by Hirono *et al*. (1984) consists of 7 contacts less than 4.0Å between the enzyme and the side-chains of Ile49I and Leu51I. There are no intermolecular hydrogen bonds in this region.

The hydrogen-bonding interactions between subtilisins and their protein inhibitors are somewhat different than those of the chymotrypsin-like enzymes. Both families of proteinases have a short segment of extended chain at the bottom of the active site cleft that forms a short β-sheet with the P_1 to P_3 residues of the inhibitor. In the subtilisin Novo:CI-2 complex, there is a β-bridge of two antiparallel hydrogen bonds between Val57I and Gly127, plus the long hydrogen bond from Met59I to Ser125 (Table 2.12). The subtilisins have an additional segment of extended chain along one side of the wall of the active site cleft that forms a short β-sheet with the P_2 to P_4 residues of the inhibitor (Hirono *et al.*, 1984). In subtilisin Novo, residues Gly100 and Gly102 form an antiparallel β-sheet with Ile56I and Thr58I of CI-2. The first half of the reactive site loop is thus the central strand of a short 3-stranded antiparallel sheet (Fig. 2.19).

Table 2.12

Hydrogen Bonds Between Subtilisin Novo and CI-2

Site	Residues	N···O(Å)	H···O(Å)	N-Ĥ-O(°)
P_4	Ile56I NH→Gly102 CO	2.87	1.89	164
	Gly102 NH→Ile56I CO	2.94	2.07	144
P_3	Val57I NH→Gly127 CO	3.01	2.01	176
	Gly127 NH→Val57I CO	2.97	2.05	152
P_2	Thr58I NH→Gly100 CO	3.15	2.28	145
P_1	Met59I NH→Ser125 CO	3.38	2.49	148
	Asn155 $N^{\delta 2}$→Met59I CO	2.73	1.77	160
	Ser221 NH→Met59I CO	3.06	2.19	144
P_2'	Tyr61I NH→Asn218 CO	2.74	1.76	167

The carbonyl oxygen of the P_1 residue, Met59I, is the recipient of two good hydrogen bonds from Ser221 and Asn155, the residues comprising the oxyanion hole (Table 2.12). The hydrogen-bonding interactions on the C-terminal side of the scissile bond are minimal. The side-chain of Asn155 is within contact distance of the NH and CO of Glu60I, but the distances are too long to be considered hydrogen bonds (3.5 and 3.8Å). The NH of Tyr61I (P_2') forms a hydrogen bond to the CO of Asn218; this interaction is conserved in the chymotrypsin-like enzymes, with a bond from P_2' to the CO of residue 41 (numbering of chymotrypsinogen A (Hartley & Kauffman, 1966)).

2.4.2 The Reactive Site Bond and the Active Site

Information on binding constants and kinetic parameters for the inhibition of subtilisin Novo by CI-2 are not available and thus whether this inhibitor and other members of its family actually act by the standard mechanism of Laskowski & Kato (1980) is not known definitively. CI-2 does bind to subtilisin Novo in the manner of a good substrate and in a similar manner to the binding of protein inhibitors that are known to act by the standard mechanism (Fujinaga *et al.*, 1982; Read *et al.*, 1983; Hirono *et al.*, 1984; Marquart *et al.*, 1983). The large number of favorable interactions between the reactive site loop of the inhibitor and the active site region of the enzyme are expected to result in a high equilibrium constant for the association of the two molecules, one of the characteristics of the standard mechanism.

Because inhibitors acting by the standard mechanism are hydrolyzed extremely slowly, the X-ray structures of their complexes show that the scissile bond is still intact. The state of this bond in CI-2 was checked at the completion of the refinement by calculating structure factors for the complex omitting all atoms of Met59I and Glu60I. These structure factors were used to calculate a difference electron density map with coefficients $m|\mathbf{F}_o|-|\mathbf{F}_c|$, α_c, and the map was examined in the region around the active site. The only significant feature seen in the map was strong continuous electron density for the two inhibitor residues that had been omitted from the calculation. This region of the map is shown in Figure 2.20. This experiment shows that the reactive site bond of CI-2 is likely uncleaved in the complex, further evidence that the inhibitor may act by the standard mechanism.

Initial unrefined structures of pancreatic trypsin inhibitor (PTI) with bovine trypsin (Rühlmann *et al.*, 1973) and soybean trypsin inhibitor (STI) with porcine trypsin (Sweet *et al.*, 1974) were interpreted as having the inhibitor covalently bound to the enzyme as a tetrahedral adduct. More recent refined structures of the trypsin:PTI system (Marquart *et al.*, 1983) and the ovomucoid inhibitors (Read *et al.*, 1983) have indicated that the interaction between the active site serine O^{γ} and the carbonyl carbon of

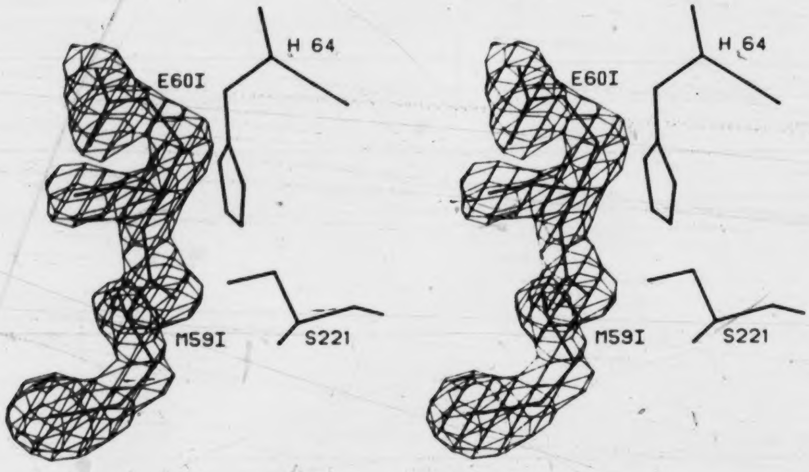


Figure 2.20. Electron Density of the Scissile Bond in CI-2. A difference electron density map of the immediate region of the active site residues, calculated omitting atoms of the P_1 and P_1' residues of CI-2. The electron density of the scissile bond is strong and continuous. Contour surfaces are drawn at 0.42 e/Å³.

the P_1 residue is too long to be a covalent bond (average length of about 2.7Å) but too short for a normal van der Waals contact (3.1Å). NMR data have also been interpreted as showing that no tetrahedral adduct is formed (Baillargeon *et al.*, 1980; Richarz *et al.*, 1980), but the results do not show whether the peptide of the scissile bond is planar or distorted to some extent toward the tetrahedral. Some workers have found distortions toward the tetrahedral in their inhibitor structures of -34° in the θ' angle (the out-of-plane bend angle of the carbonyl oxygen relative to C^{α}, N and the carbonyl carbon) (Huber & Bode, 1978), and others have found distortions of only -5° (Fujinaga *et al.*, 1982; Read *et al.*, 1983). The current concensus seems to be that some distortion is observed when several independent structures are compared (Marquart *et al.*, 1983), but that the amount of distortion is at the limits of detectability given the accuracy of the crytallographic structures (Read & James, 1986).

The peptide of the reactive site bond of CI-2 shows little if any distortion from the expected planar trigonal carbonyl carbon atom. The torsion angle ω (C^{α}(59I)-CO-N-C^{α}(60I)) is 175°, not significantly different from the expected value of 180°. The angle θ' of this bond is 1°, indicating that no pyramidalization has occurred in the complex. The distance from O^{γ} of Ser221 to the carbonyl carbon of Met59I (P_1) is 2.73Å, similar to the equivalent distance in other serine proteinase:inhibitor complexes (Read *et al.*, 1983).

The geometry of His64 and Ser221 in the active site of the enzyme in relation to the scissile bond is generally very similar to that of the equivalent residues in the chymotrypsin-like enzymes (Fujinaga *et al.*, 1982). The hydrogen bond from His64 $N^{\epsilon 2}$ to Ser221 O^{γ} is 2.72Å long. The relative orientation of Asp32 to His64 in the other interaction of the catalytic triad is quite different to that of Asp102 and His57 in, for example, the complex of OMTKY3 with SGPB (Fujinaga *et al.*, 1982). In the chymotrypsin-like enzymes, the plane of the Asp carboxylate group is at an angle of -30° to the plane of the His imidazole ring. The carboxylate group of the Asp receives 4 hydrogen bonds from His57 $N^{\delta 1}$, Ser214 O^{γ}, His57 NH, and Gly56 NH. In contrast, in the subtilisin Novo:CI-2 complex, the carboxylate group of Asp32 is at an angle of +27° with the imidazole ring of His64 (Fig. 2.19). His64 $N^{\delta 1}$ provides a bifurcated hydrogen bond to both the oxygen atoms of the carboxylate group, and the two other hydrogen bonds to the carboxylate are provided by Ser33 NH and a water molecule buried behind the active site. A more detailed description of differences in the active site regions of the subtilisins and the chymotrypsin-like enzymes is given in chapter 5.

One other feature seen in subtilisin Novo that may interact with an inhibitor or substrate is the helix dipole of helix F. This helix begins at Gly219 with three type III 3_{10} turns; the regular helix begins at Ser224 and is 3.9 turns long. The positively charged end of the helix dipole points at Ser 221, and the interaction of the two may enhance the catalytic action of subtilisin by promoting proton transfer (Hol, 1985). The positive charge of this dipole is also well-positioned to add to the polarizing environment of the oxyanion hole, which would promote proton transfer and a nucleophilic attack of the Ser221 O^{γ} on the P_1 residue, as discussed in section 1.3.

2.4.3 Solvent Structure of the Complex

The final refined model of the subtilisin Novo:CI-2 complex includes 167 solvent molecules; 165 of these have been treated as water molecules and the remaining two have been refined as Ca^{2+} ions (section 2.2.2). Rough calculations of the amount of solvent to be expected in the asymmetric unit of these crystals indicate that only 10 to 15% of the solvent present has been modelled. This low percentage is due in part to the conservative criteria used for choosing water molecules and for retaining less ordered solvent in the model (section 2.1.4). In structures of proteins of a similar size and with a similar solvent content in the crystals, 2 or 3 times as many water molecules can be reliably defined, although the occupancies of the weaker ones may range as low as 0.3 (James & Sielecki, 1983).

In regions of close contact between proteins in a crystal, deciding which molecule that a solvent atom 'belongs to' can be difficult. To be sure of including all possible solvent atoms from the refined model that are interacting with subtilisin Novo or CI-2, symmetry-related solvent atoms within 4.0Å of a protein atom were added to the model for the following analysis. A total of 193 waters were thus part of all calculations. The complete solvent structure of the subtilisin Novo:CI-2 complex is shown in Figure 2.21.

The distribution of water molecules on the surface of the complex was examined by comparing the accessibility of surface residues calculated in the presence and absence of

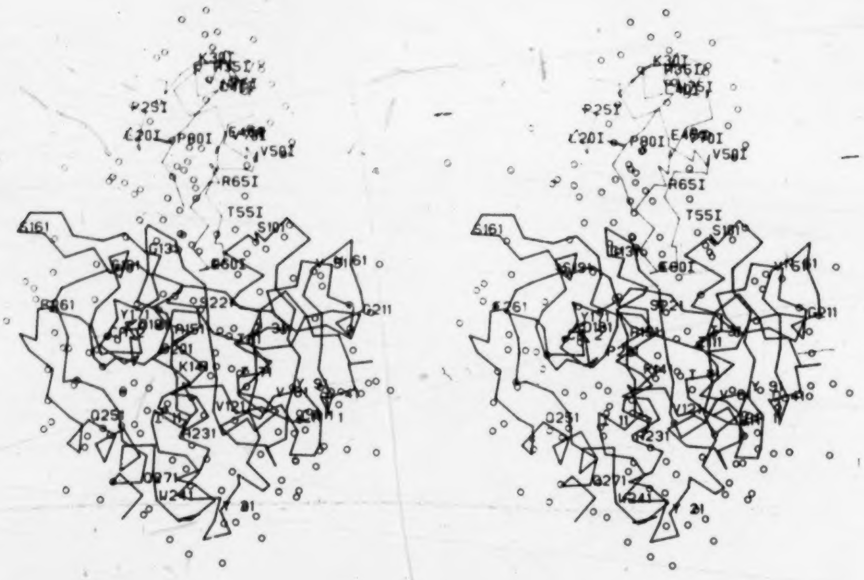
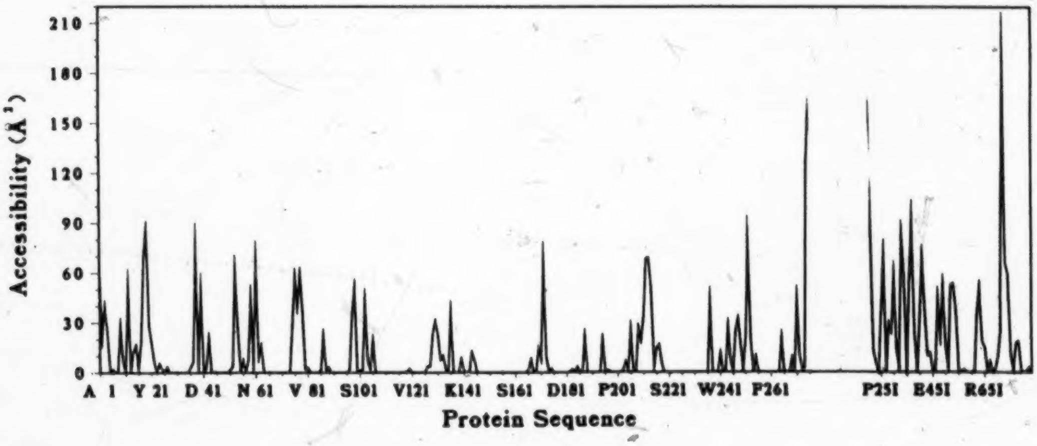


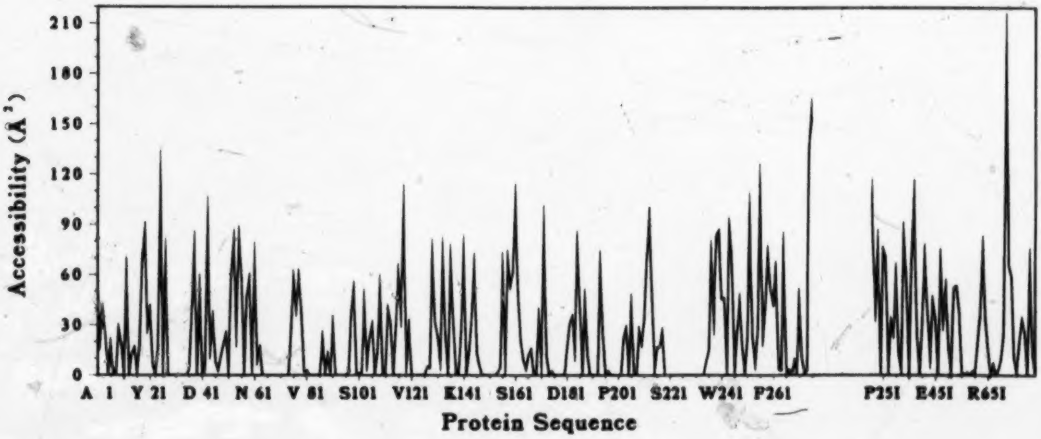
Figure 2.21. Solvent Molecules Associated With the Complex. Water molecules are indicated by open circles, Ca^{2+} ions by closed circles, the main-chain atoms of subtilisin Novo with thick lines, and the main-chain atoms of CI-2 with thin lines. All solvent molecules in the model of the crystal structure and all symmetry-related solvent molecules within 4.0Å of a protein atom have been included in this figure.

water. Accessibility of all the atoms in the enzyme:inhibitor complex was calculated with the algorithm of Lee & Richards (1971), using a local program written by J. Moult. Four accessibility calculations were made: the protein complex with all solvent present plus all neighboring symmetry-related solvent and protein atoms, the protein complex without solvent but with symmetry-related protein atoms, the protein complex with solvent but without symmetry-related protein atoms, and the protein complex alone (Fig. 2.22). Atoms were considered to be accessible (and thus presumably at the protein surface) if more than 10% of their total surface area was accessible in the

a


Accessibility (Å²)
210
180
150
120
90
60
30
0
A 1 Y 21 D 41 N 61 V 81 S101 V121 K141 S161 D181 P201 S221 W241 P261 P251 E451 R651
Protein Sequence

b


Accessibility (Å²)
210
180
150
120
90
60
30
0
A 1 Y 21 D 41 N 61 V 81 S101 V121 K141 S161 D181 P201 S221 W241 P261 P251 E451 R651
Protein Sequence

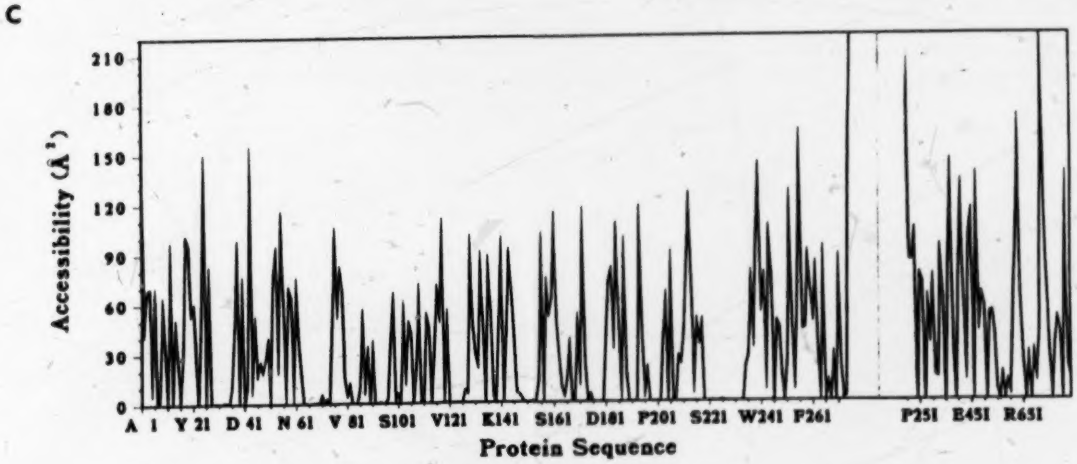


Figure 2.22. Residue Accessibilities in the Complex. Accessibilities are calculated for each atom in the complex by the method of Lee & Richards (1971). Residue accessibilities are the sum of atomic acessibilities, in Å². a) Accessibility in the presence of all solvent and symmetry-related molecules. b) Accessibility with symmetry-related protein molecules removed but all solvent within 4.0Å of the protein surface retained, including symmetry-related solvent. c) Accessibility in the absence of all solvent and symmetry-related molecules.

absence of solvent or symmetry-related protein molecules. By this criterion, 15% (367) of the protein atoms in the complex are accessible; this is similar to the percentage seen for another protein of comparable size analyzed by this method (*Streptomyces griseus* trypsin, R. Read & J. Moult, unpublished results).

A total of 232 of the accessible atoms defined by the above procedure show little change in their accessibility in the presence or absence of solvent molecules when symmetry-related molecules are present. These atoms tend to be clustered in regions on the surface of the complex and these clusters can be termed 'bare patches', i.e. regions that are lacking in coverage by ordered solvent molecules. About 65% of the atoms in the bare patches lie in regions of contact with symmetry-related molecules. This can be deduced from the increase in the calculated accessibility of these atoms in the absence of symmetry-related molecules compared to accessibility in their presence (Fig. 2.22).

The remaining 35% of the atoms in bare patches can be considered truly exposed on the surface of the complex in the crystals. All of these atoms belong to residues in turns or segments of irregular structure on the complex surface and most of the residues have hydrophilic or charged side-chains. Approximately half of these true bare patches are regions of above-average temperature factor; because the protein residues are more disordered in these areas, solvent molecules associated with them may be more disordered also and probably would not meet the criteria used for solvent selection. It is difficult to see why the remaining bare patches have little associated solvent, since they have moderate or low *B*-factors relative to the overall average of the complex.

Water molecules in the subtilisin Novo:CI-2 complex make a total of 210 hydrogen bonds with polar protein oxygen and nitrogen atoms, and an additional 101 hydrogen bonds are made between water molecules. The geometric criteria for choosing these hydrogen bonds are the same as those used for defining hydrogen bonds in secondary structural elements (section 2.2.1). The hydrogen bonds between water and protein atoms can be broken down into 4 groups; 18% are between water and main-chain nitrogen atoms, 47% between water and main-chain oxygen atoms, 8% between water and side-chain nitrogen atoms, and 27% between water and side-chain oxygen atoms.

Of the 193 solvent molecules within 4.0Å of the surface of the complex, 33 (17%) can be classified as buried in the interior of the enzyme. Two of these are the Ca^{2+} ions; Ca^{2+} ion 1 is completely buried by the residues coordinating to it; Ca^{2+} ion 2 is completely buried in the complex with solvent included, and 3% accessible when other solvent molecules are omitted from the calculation. Nineteen of the buried water molecules have accessibilities of less than 5% when other solvent and symmetry-related molecules are omitted from the calculation, and the remaining 12 have accessibilities between 5 and 10%.

Several of the buried water molecules are located close to the active site. The buried polar side-chain of His67 lies close behind the active site His64 and Ser221 and is far from bulk solvent in the complex, but it is still connected to bulk solvent through a long channel of well-ordered water molecules that leads to the surface of the enzyme near the N-terminus (Fig. 2.23). This channel is composed of O344, between the side-chains of Met222 and His67, Thr71 O^{γ}, O360, O367, O398, and O384 at the bottom of a surface cleft. The buried side-chain of His226 is also connected to this channel via O353→O360 and O367. Another

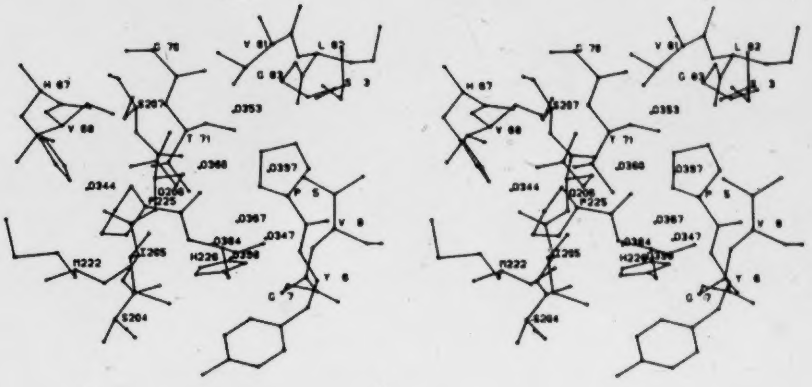


Figure 2.23. Buried Water Channel in Subtilisin Novo. The water molecules forming a channel from the buried side-chains of His64, Thr71, and His226 to the enzyme surface, with nearby protein residues. The channel stretches from behind the active site, under the 'arm' of antiparallel β-sheet straddling the two halves of the enzyme, and emerges close to the N-terminus.

important buried water molecule at the active site is O370, which bridges the center of the 5-membered turn containing Asp32. O352 provides one of the four stabilizing hydrogen bonds to the side-chain of Asp32 and also has hydrogen bonds with the NH of Ser125 and the $N^{\delta2}$ of Asn123.

One water molecule associated with the inhibitor is buried, O374. This is one of the chain of water molecules that lies between strand 2 of the parallel β-sheet and the C-terminal strand, providing hydrogen bonds to both segments of protein. Only two water molecules are trapped between the enzyme and the inhibitor in the active site region and thus rendered inaccessible. O372 lies at the bottom of the P_1 specificity pocket, hydrogen bonding with the carbonyl oxygen atoms of Leu126 and Ala152, and the NH of Gly169. O396 is between Thr55I and Pro129, at the outer end of the short 3-stranded β-sheet formed by the inhibitor and the enzyme.

Some of the buried water molecules lie close to the surface of the complex in clefts that are covered by the side-chain of a single residue. The covering residue is often one with a long flexible side-chain, thus these water molecules may have more access to bulk solvent than is indicated by a static measure of their accessibility. There are an additional 32 solvent molecules that are accessible in the absence of other solvent, but can be considered buried in its presence. A number of these lie at the bottom of clefts on the complex surface. Water molecules with the same accessibility in the presence or absence of other solvent, but with increased accessibility in the absence of symmetry-related molecules are probably buried in the interfaces between the symmetry-related molecules. There are 43 water molecules of this type. The total number of water molecules that may be considered buried, either inside the complex, in clefts beneath other ordered water, or between symmetry-related molecules, is 108. This is 56% of all the water associated with the complex.

2.4.4 Accessibility of Protein Residues

The variation in accessibility of individual protein residues along the polypeptide chains of subtilisin Novo and CI-2 is shown in Figure 2.21. The differences in accessibility in the presence of solvent or symmetry-related molecules have been discussed in the previous section. Because it is a much smaller protein, the accessibility of CI-2 varies much more rapidly along the chain than the accessibility of subtilisin Novo. Residues in the active site region and the reactive site loop are almost totally inaccessible, even in the absence of solvent and symmetry-related molecules (Fig. 2.21c).

Table 2.13 gives a summary of accessibilities for each residue type in the complex, both in the crystal environment with solvent and symmetry-related molecules present and in their absence. The percentage of each residue type buried in this complex structure can be compared to the percentages found by Chothia (1975) in a survey of 9 globular proteins. Chothia calculated potential accessible surface area by the method of Lee & Richards (1971), and used fully extended side-chains in a Gly-X-Gly peptide as standard residue conformations. With this procedure, Chothia obtained numbers for potential accessible surface area for each residue similar to those of Shrake & Rupley (1973); thus Chothia's percentages of buried residues of each type can be compared to the percentages for the subtilisin Novo:CI-2 complex in the absence of solvent and symmetry-related molecules.

Table 2.13

Accessibilities by Residue Type

Residue	No.	% Buried[1]		% Partially buried[2]		% Exposed		% Buried (Chothia)[3]
		+[4]	-	+	-	+	-	
Ile	19	79	68	0	0	21	32	65
Val	40	85	70	5	8	10	22	56
Met	6	100	67	0	17	0	16	50
Phe	4	100	0	0	25	0	75	48
Cys	0	-	-	-	-	-	-	47
Leu	21	76	57	0	10	24	33	41
Ala	39	82	62	5	8	13	30	38
Gly	35	86	51	0	11	14	38	38
Thr	16	75	38	12	12	13	50	25
Ser	38	55	21	11	0	34	79	24
Trp	4	75	25	25	0	0	75	23
Pro	18	56	22	11	17	33	61	21
Glu	11	45	9	0	0	55	91	20
His	6	83	67	17	0	0	33	19
Asp	14	57	14	7	14	36	72	15
Tyr	11	73	0	18	18	9	82	13
Asn	19	58	5	16	5	26	90	10
Gln	12	17	0	0	0	83	100	6
Lys	16	31	6	12	0	57	94	4
Arg	6	50	0	17	17	33	83	0

[1] A residue is considered buried if the sum of its atomic accessibilities is less than 5% of its potential accessible surface area in a fully extended conformation as the central residue of a Gly-X-Gly peptide (Shrake & Rupley, 1973). N- and C-terminal residues were excluded from these calculations.

[2] A residue is considered partially buried if the sum of its atomic accessibilities is between 5 and 10% of its potential accessible surface area.

[3] From a survey of 9 globular proteins (Chothia, 1975).

[4] 'Plus' is accessibility calculated with all solvent and symmetry-related molecules present, 'minus' is with no solvent or symmetry-related molecules.

The majority of the residue types in the subtilisin Novo:CI-2 complex have a similar percentage of buried residues to the proteins in Chothia's survey, within 10 to 15%. Some of the larger differences, such as that for phenylalanine residues, are probably due to the small number of that particular residue type in the complex. The most striking difference is for histidine residues, with 4 out of 6 being buried in this complex; the residues considered buried are His39, His64, His67, and His226. His39 is sandwiched between two loops in subtilisin Novo, close to the enzyme surface. Its side-chain forms hydrogen bonds with the carbonyl oxygen of Lys213 and the amide nitrogen of Asp41. His64 is the catalytic histidine and is, of course, buried at the interface between the enzyme and the inhibitor. His67 lies behind the active site in subtilisin Novo, some distance from the enzyme surface in any direction. It forms hydrogen bonds to members of a channel of buried water molecules that leads to the enzyme surface near the N-terminus. His226 is close to His67 and interacts with the same channel of buried water molecules. This channel of water molecules is buried by the 'arm' of antiparallel β-sheet that reaches across from domain 1 to domain 2. All of the buried histidines thus have fully hydrogen bonded side-chains and, if they are charged at the pH of their micro-environment, may balance that charge through indirect contact with bulk solvent. The nature of their hydrogen bonds indicates, however, that all but His64 are neutral in their protein environment.

Eight other charged residues in the complex are either buried (less than 5% exposed) or partially buried (less than 10% exposed). They include Asp32, in the active site, and

Glu60I, Arg65I, and Gly83I in the reactive site loop of the inhibitor. These 3 residues are fully involved in hydrogen-bonding or electrostatic interactions in the active site region (section 2.4.2). Asp36 lies in a surface cleft and its side-chain participates in 4 hydrogen bonds, one of them to an ordered water molecule on the enzyme surface. Asp41 is the single negatively charged ligand for Ca^{2+} ion 1. The side-chain of Asp60 participates in 4 hydrogen bonds to protein ligands and is also close to the N-terminal end of helix C; the negative charge of the Asp is at least partly compensated by the positive helix dipole. Lys94 is fully buried and forms a salt bridge with Glu54.

Chothia (1975) has shown that the molecular weight of a globular protein is proportional to the accessible surface area that protein would have in the unfolded and fully extended conformation.

$$A_T = 1.44 \times M$$

where M is molecular weight. The proportion of the total accessible surface that is buried on folding the protein is linearly related to molecular weight, and the acessible area buried on folding is

$$A_B = (0.859 \times M) + (0.774 \times M^2 \times 10^{-5})$$

The accessible surface area of the native protein is the difference between A_T and A_B:

$$A_S = (0.581 \times M) - (0.774 \times M^2 \times 10^{-5})$$

The accessible surface area of the subtilisin Novo:CI-2 complex from this calculation, assuming a complex molecular weight of 36863, is 11300Å^2. The accessible surface area calculated as the sum of the accessible atomic surface areas obtained from the program of J. Moult, is 12100Å^2. These two values are reasonably close; the formulae of Chothia are applicable strictly only to monomeric proteins.

The accessible surface areas of the enzyme and inhibitor can be calculated in the absence of the other member of the complex. The sum of the accessible atomic surface areas for subtilisin Novo in the absence of CI-2 is 9300Å^2, and that for CI-2 in the absence of subtilisin Novo is 4400Å^2. The value for CI-2 is much higher than would be predicted from its molecular weight, but it is a much smaller and less globular protein than those used to derive the proportionality constants of Chothia. The difference in accessible surface area between the complex and the sum of the two components of the complex is 1600Å^2. Wodak & Janin (1978) have calculated the amount of accessible surface area buried on the binding of the inhibitor PTI to trypsin to be between 1300 and 1500Å^2. Assuming a free energy gain of 24 cal. for each Å^2 removed from contact with water on formation of the complex (Chothia, 1975; Lee & Richards, 1971), the free energy gain on binding of CI-2 to subtilisin Novo is 38.4 Kcal/mole. This energy gain can be translated to a huge and obviously unrealistic value of 10^{28} for the binding constant. Other factors are important; the free energy of association calculated from the amount of hydrophobic surface buried is a large part of the total free energy of association, but contributions from polar and van der Waals interactions must be included to be able translate the free energy to a binding constant (Wodak & Janin, 1978). The formation of a complex from two molecules also adds an unfavorable entropic term to the calculation, especially if there is a loss of internal degrees of motional freedom for the inhibitor on binding (see section 5.6).

Bibliography

Alden, R. A., Birktoft, J. J., Kraut, J., Robertus, J. D., & Wright, C. S. (1971) *Biochem. Biophys. Res. Commun. 45*, 337-344.

Arnott, S. & Dover, S. D. (1967) *J. Mol. Biol. 30*, 209-212.

Artymiuk, P. J. & Blake, C. C. F. (1981) *J. Mol. Biol. 152*, 737-762.

Baillargeon, M. W., Laskowski, M. Jr., Neves, D. E., Porubcan, M. A., Santini, R. E., & Markley, J. L. (1980) *Biochemistry 19*, 5703-5710.

Baker, E. N. (1980) *J. Mol. Biol. 141*, 442-484.

Baker, E. N. & Hubbard, R. E. (1984) *Prog. Biophys. Molec. Biol. 44*, 97-179.

Barry, C. D., Molnar, C. E., & Rosenberger, F. U. (1976) *Technical Memo No. 229*, Computer Systems Lab, Washington University, St. Louis, MO.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. J. Jr., Brice, M. D., Rogers, J. K., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535-542.

Brown, I. D. & Shannon, R. D. (1973) *Acta Cryst. A29*, 266-282.

Chothia, C. (1975) *Nature 254*, 304-308.

Chothia, C., Levitt, M., & Richardson, D. (1981) *J. Mol. Biol. 145*, 215-250.

Crawford, J. L., Lipscomb, W. N., & Schellman, C. G. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 538-542.

Crowther, A. R. (1973) in *The Molecular Replacement Method*, International Science Review *13*, ed. M. G. Rossmann, New York: Gordon & Breach, 173-178.

Cruickshank, D. W. J. (1949) *Acta Cryst. 2*, 65-82.

Cruickshank, D. W. J. (1954) *Acta Cryst. 7*, 519.

Cruickshank, D. W. J. (1967) in *International Tables for X-ray Crystallography* Vol. 2, eds. J. S. Kasper & K. Lonsdale, Birmingham: Kynoch Press, 318-340.

Davies, D. R. & Segal, D. M. (1971) in *Meth. Enzymol.*

Vol.XXII, ed. W. B. Jacoby, New York: Academic Press, 266-269.

Drenth, J., Hol, W. G. J., Jansonius, J. N., & Koekoek, R. (1972) *Eur. J. Biochem. 26*, 177-181.

Einspahr, H. & Bugg, C. E. (1981) *Acta Cryst. B37*, 1044-1052.

Einspahr, H. & Bugg, C. E. (1984) in *Metal Ions in Biological Systems* Vol. 17, ed. H. Sigel, New York: Marcel Dekker, 51-97.

Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M. Jr., & James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868-4872.

Hartley, B. S. & Kauffman, D. L. (1966) *Biochem. J. 101*, 229-231.

Hendrickson, W. A. (1976) *J. Mol. Biol. 106*, 889-893.

Hendrickson, W. A. (1981) in *Refinement of Protein Structures, Proc. Daresbury Study Weekend, 15-16 Nov. 1980*, eds. P. A. Machin, J. W. Campbell, & M. Elder, Warrington: Daresbury Laboratory, 1-8.

Hendrickson, W. A. & Konnert, J. H. (1980) in *Biomolecular Structure, Function, Conformation and Evolution* Vol. I, ed. R. Srinivasan, Oxford: Pergamon Press, 43-57.

Hirono, S., Akagawa, H., Mitsui, Y., & Iitaka, Y. (1984) *J. Mol. Biol. 178*, 389-413.

Hol, W. G. J. (1985) *Prog. Biophys. Molec. Biol. 45*, 149-195.

Huber, R. & Bode, W. (1978) *Acc. Chem. Res. 11*, 114-122.

James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol. 163*, 299-361.

Johnson, C. K. (1965) ORTEP, Report ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, TN.

Jonassen, I. (1980) *Carlsberg Res. Commun. 45*, 47-58.

Kabsch, W. & Sander, C. (1983) *Biopolymers 22*, 2577-2637.

Kretsinger, R. H. (1976) *Ann. Rev. Biochem. 45*, 239-266.

Kraut, J. (1977) *Ann. Rev. Biochem. 46*, 331-358.

Laskowski, M. Jr. & Kato, I. (1980) *Ann. Rev. Biochem. 49*, 593-626.

Lee, B. & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379-400.

Lifchitz, A. (1983) *Acta Cryst. A39*, 130-139.

Luzzati, V. (1952) *Acta Cryst. 5*, 802-810.

Luzzati, V. (1953) *Acta Cryst. 6*, 142-152.

Markland, F. S. & Smith, E. (1971) in *The Enzymes, 3rd ed.* Vol. 3, ed. P. Boyer, New York: Academic Press, 561-608.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Cryst. B39*, 480-490.

Matsubara, H., Hagihara, B., Nakai, M., Komaki, T., Yonetani, T., & Okunuki, K. (1958) *J. Biochem. (Tokyo) 45*, 251-255.

Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491-497.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., & Kraut, J. (1977) *J. Biol. Chem. 252*, 8875-8883.

McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, New York: John Wiley & Sons, 96-97.

North, A. C. T., Phillips, D. C., & Mathews, F. S. (1968) *Acta Cryst. A24*, 878-884.

Olaitan, S. A., DeLange, R. J., & Smith, E. L. (1968) *J. Biol. Chem. 243*, 5296-5301.

Ottesen, M. & Spector, A. (1960) *Compt. Rend. Trav. Lab. Carlsberg 32*, 63-74.

Patterson, A. L. (1934) *Phys. Rev. 46*, 372-376.

Ramakrishnan, C. & Ramachandran, G. N. (1965) *Biophys. J. 5*, 909-933.

Rao, S. N., Jih, J.-H., & Hartsuck, J. A. (1980) *Acta Cryst. A36*, 878-884.

Read, R. J. (1986a) Ph. D. Thesis, University of Alberta.

Read, R. J. (1986b) *Acta Cryst.*, in press.

Read, R. J. & James, M. N. G. (1986) in *Proteinase Inhibitors*, ed. A .J .Barrett, Amsterdam: Elsevier, in press.

Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420-4433.

Read, R., Fujinaga, M., Sielecki, A., Ardelt, W., Laskowski, M. Jr., & James, M. (1984) *Acta Cryst. A40*, C50-C51.

Richardson, J. S. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2619-2623.

Richardson, J. S. (1981) *Adv. Prot. Chem. 34*, 167-339.

Richarz, R., Tschesche, H., & Wüthrich, K. (1980) *Biochemistry 19*, 5711-5715.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., & Wilcox, P. E. (1972a) *Biochemistry 11*, 2439-2449.

Robertus, J. D., Kraut, J., Alden, R. A., & Birktoft, J. J. (1972b) *Biochemistry 11*, 4293-4303.

Rossmann, M. G. (1973) *The Molecular Replacement Method*, International Science Review 13, New York: Gordon & Breach.

Rossmann, M. G. & Blow, D. M. (1961) *Acta Cryst. 14*, 641-647.

Rühlmann, A., Kukla, D., Schwager, P., Bartels, K., & Huber, R. (1973) *J. Mol. Biol. 77*, 417-436.

Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157-162.

Seemüller, U., Eulitz, M., Fritz, H., & Strobl, A. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1841-1846.

Shrake, A. & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351-372.

Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1979) *J. Mol. Biol. 134*, 781-804.

Sielecki, A. R., James, M. N. G., & Broughton, C. G. (1982) in *Computational Crystallography*, ed. D. Sayre, Oxford: Oxford University Press, 409-419.

Smith, E. L., DeLange, R. J., Evans, H., Landon, M., & Markland, F. S. (1968) *J. Biol. Chem. 243*, 2184-2191.

Svendsen, I., Jonassen, I., Hejgaard, J., & Boisen, S. (1980b) *Carlsberg Res. Commun. 45*, 389-395.

Svendsen, I., Martin, B., & Jonassen, I. (1980a) *Carlsberg Res. Commun. 45*, 79-85.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., & Blow, D. M. (1974) *Biochemistry 13*, 4212-4228.

Thiessen, W. E. & Levy, H. A. (1973) *J. Appl. Cryst. 6*, 309.

Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., & Chen, E. Y. (1983) *Nucleic Acids Research 11*, 7911-7925.

Wodak, S. J. & Janin, J. (1978) *J. Mol. Biol. 124*, 323-342.

Wright, C. S., Alden, R. A., & Kraut, J. (1969) *Nature (London) 221*, 235-242.

3. Subtilisin Carlsberg and Eglin-c[1]

Subtilisin Carlsberg was the first bacterial serine proteinase to be discovered (Linderström-Lang & Ottesen, 1947). It consists of a single polypeptide chain of 274 amino acids, M_r = 27292 (Smith *et al.*, 1968), and has strong sequence homology with subtilisin Novo (≡ BPN', Fig. 2.1). The catalytic activity and substrate specificity of subtilisin Carlsberg are qualitatively similar to those of subtilisin Novo but subtilisin Carlsberg has a higher ratio of esterase to proteinase activity (Barel & Glazer, 1968). The two enzymes do not cross-react serologically (Keay & Moser, 1969), despite being 70% identical in amino acid sequence.

Eglin-c is a small protein inhibitor of serine proteinases isolated from the leech *Hirudo medicinalis* (70 amino acids, M_r = 8092) (Seemüller *et al.*, 1977; Seemüller *et al.*, 1980) (Fig. 2.2). It is a member of the potato inhibitor 1 family of inhibitors, most of whose members lack stabilizing disulphide bridges (Melville & Ryan, 1972; Svendsen *et al.*, 1982). Eglin-c is a very strong inhibitor of mammalian leukocyte elastase, cathepsin G and α-chymotrypsin ($K_i \simeq 10^{-11}$ M), but not of other mammalian proteinases tested (Schnebli *et al.*, 1985). It is also a good inhibitor of several subtilisins. In the kinetic studies and the present structure determination, the genetically engineered product, *N*-acetyleglin-c (M_r = 8111), was used (Rink *et al.*, 1984).

---

[1] A version of portions of this chapter has been published [McPhalen, C. A., Schnebli, H. P., & James, M. N. G. (1985) *FEBS Letters 188*, 55-58].

Crystals of a molecular complex of subtilisin Carlsberg and eglin-c have been grown. Grütter *et al.* (1985) have reported crystals of a molecular complex of subtilisin Carlsberg and eglin-c that are isomorphous with the subtilisin Carlsberg:eglin-c crystals studied here, but grown under different crystallization conditions. The molecular replacement method (Rossmann, 1973) was used to obtain an initial electron density map for the complex. The majority of the polypeptide backbone of the inhibitor could be fitted to this map, as well as the complete sequence of the enzyme. The least-squares refinement of the structure was completed after 57 cycles, at an *R*-factor of 0.136 for data in the resolution range 8.0 to 1.8Å. The secondary and tertiary structures of the enzyme and inhibitor have been examined in some detail, as well as the interactions between the two in the complex. The subtilisin Carlsberg:eglin-c complex is compared to the subtilisin Novo:CI-2 complex in chapter 5.

3.1 Structure Solution and Refinement

3.1.1 Crystallization

Purified lyophilized eglin-c (batch #84, 25-4) was obtained from Ciba-Geigy AG., Basel, Switzerland (Rink *et al.*, 1984). Subtilisin Carlsberg was the kind gift of Dr. I. Svendsen, Carlsberg Laboratories, Copenhagen, Denmark. Small rod-shaped crystals of the complex were grown by the hanging-drop vapor-diffusion method (Davies & Segal, 1971; McPherson, 1982). Subtilisin Carlsberg (1.7 mg/mL) and eglin-c (0.6 mg/mL, a 1.2 M excess of eglin-c to subtilisin Carlsberg) were dissolved in distilled water and a hanging drop was prepared consisting of equal amounts of the protein solution and a buffer of 0.7 M KH_2PO_4 and K_2HPO_4 at pH 5.6. The reservoir of the crystallization tray was filled with 1.0 mL of buffer solution. The crystals grew to their maximum size in less than a week, but were usually small for high resolution data collection purposes. Small seed crystals were placed into freshly prepared hanging drops of protein and buffer solution, and grew within a week or two to an average size of 0.8x0.4x0.2 mm.

Analysis of precession photographs showed that the crystals diffract to a minimum d spacing of 1.8Å. No systematic absences were observed; the lattice symmetry is consistent with triclinic space group P1. The unit cell dimensions of the crystals are a = 38.31(3)Å, b = 41.41(4)Å, c = 56.50(6)Å, α = 69.5(1)°, β = 83.7(2)°, and γ = 75.3(1)°. These dimensions are the same as those found by Grütter *et al.* (1985) if the different conventions used to assign the unit cell angles are taken into account. The volume per unit molecular weight of the crystals, V_m, is 2.29Å^3/dalton, consistent with one molecule of the complex (M_r = 35384) per unit cell and a solvent content of 46% (Matthews, 1968).

3.1.2 Data Collection and Processing

Intensity data sets were collected from two crystals, described in Table 3.1. Both data sets were collected on a CAD4 diffractometer. The incident radiation was Ni-filtered CuK_α, at 40 kV and 26 mA, using a 1.3 mm diameter incident beam collimator. The crystal-to-counter distance was 60 cm, and the diffracted beam passed through a He-filled tunnel. The peaks were scanned 0.8° in ω at 0.67°/min., and backgrounds were scanned 0.2° in ω on either side of the peak scan. To minimize loss of weak reflections at higher resolution due to decay, both data sets were collected in shells. Three shells were collected for the 2.0Å data set; first from 2.0 to 3.0Å, then from 3.0 to 5.0Å, and finally from 5.0 to 37.9Å. Five shells were collected for the 1.8Å data set; first from 1.8 to 1.9Å, then 1.9 to 2.1Å, 2.1 to 2.5Å, 2.5 to 3.2Å, and finally from 3.2 to 37.9Å.

During data processing, the measured backgrounds of the reflections collected were corrected for intensity dependence and an averaging function of 2θ and ϕ was applied to them. The absorption correction applied was an empirical function of ϕ (North *et al.*, 1968). Intensity decay due to radiation damage was monitored by periodic measurements of 5 standard reflections. The applied decay correction was a function of the degree of decay of these standards with time. Lorentz and polarization corrections were also applied to the data. The absolute scale factor and an overall temperature factor (B, Å^2) were calculated for each data

Table 3.1

Crystal Data for the Subtilisin Carlsberg:Eglin-c Complex

	Crystal 1	Crystal 2
Resolution range (Å)	37.9 - 2.0	37.9 - 1.8
Total reflections measured	23596	33153
Total unique reflections	21906	29275
Total reflections $I \geq \sigma(I)$[1]	19143	27517
R_{merge}[2]	0.062	0.038
Maximum absorption correction	1.26	1.48
Maximum decay correction	28 %	42 %
Scale	10.33	5.57
Overall B (Å^2)	10	12

[1] $\sigma(I) = (I + c^2 I^2 + (t_I/t_{Bk})^2 (\Sigma Bk + c^2 \Sigma Bk^2))^{1/2}$, where I = total intensity; Bk = background counts; t_I = time for intensity measurement; t_{Bk} = time for background measurement; c = instrument instability constant = 0.01.
[2] $R_{merge} = \Sigma_{hkl}(\Sigma_i |I_i - \langle I \rangle| / \Sigma_i I_i)$ for reflections measured more than once in a data set.

set with the program ORESTES (Thiessen & Levy, 1973).

3.1.3 Structure Determination

The method of molecular replacement was used to solve the phase problem for the structure of eglin-c in complex with subtilisin Carlsberg (Rossmann, 1973). The search model was the structure of subtilisin Novo as determined in this laboratory in complex with the inhibitor CI-2 (McPhalen *et al.*, 1985). This search model had been partially refined to an R-factor of 0.193 at 2.1Å resolution. Although the model enzyme and the enzyme in the complex are not the same, they do have 70% sequence identity, and the conserved regions between the two were believed to be in the enzyme core; thus, it was reasonable to expect the two to have similar conformations (Wright *et al.*, 1969). As in the molecular replacement solution for the subtilisin Novo:CI-2 complex, the model accounts for a high percentage of the scattering matter expected in the complex, 77%. Data from crystal 1 (Table 3.1) were used for the molecular replacement calculations. Detailed explanations of procedures and choice of parameters in the molecular replacement method followed here are given in section 2.1.3.

The rotation search was performed with the fast rotation function of Crowther (1973) (program of E. Dodson); the agreement between the normalized structure factors ($|\mathbf{E}|$'s) of the model and the subtilisin Carlsberg:eglin-c complex was maximized in the search. The $|\mathbf{E}_C|$'s were calculated for the subtilisin Novo model placed in an orthogonal unit cell of symmetry P1, with each edge of the cell measuring 65Å. The $|\mathbf{E}_O|$'s were calculated for the unknown structure of the complex, from the structure factors derived from data processing, with the program ORESTES (Thiessen & Levy, 1973). The Patterson maps calculated from these two sets of $|\mathbf{E}|$'s were compared only between 4.0 and 18.0Å to limit the comparison to intra-molecular vectors. The resolution range of reflections included in the rotation function search was 10.0 to 3.0Å. $|\mathbf{E}_C|$'s with magnitudes greater than 2.5 were used in the search; there were 2358 structure factors above this limit. 1636 $|E_o|$'s were included, with magnitudes above 1.5.

The initial coarse rotation function search was performed on an asymmetric unit of rotation function space (Rao *et al.*, 1980), along a grid of 5° in each of the Eulerian angles α, β, and γ (Crowther, 1973). The maximum peak in the rotation function map was 17.0σ above the mean and 11.7σ above the second highest peak. This result indicates the correct solution to the rotation problem quite unambiguously, according to previous experience in this lab. A second rotation search was performed with 1° intervals in β and 5° intervals in α and γ about the maximum peak from the coarse rotation search. This fine search gave a maximum peak of 16.4σ above the mean of the second map. The final values of the orientation parameters from the rotation search were interpolated from the fine search map to 0.2° in β and 1° in α and γ.

No translation search was required to position the subtilisin Carlsberg:eglin-c complex in its unit cell, because the space group of the complex crystals is P1 and any point in the cell may be chosen as the origin. The local translation function program, written by M. Fujinaga, may be used to refine the rotational parameters found by the rotation function search, using a 6-dimensional rotation and translation search over a small area of the translation function map. The translation search algorithm is a correlation coefficient search on $|F|^2$. The orientation found by the 6-dimensional search was very close to that found by the rotation function search; the maximum deviation between the two in any of the 3 angles was 0.37°.

The *R*-factor for the transformed coordinates of the subtilisin Novo model was 0.47 for all data to 2.0Å resolution. The initial phasing model was this subtilisin Novo molecule alone; no contribution from the inhibitor was included. The model was used in the calculation of structure factor amplitudes and phases for the computation of an initial electron density map. All electron density maps used in this structure solution were calculated with the coefficients of Read (1986), which are derived from an expression designed to suppress model bias resulting from phasing by partial structures with errors.

3.1.4 Structure Fitting and Refinement

Throughout the fitting and refinement of the subtilisin Carlsberg:eglin-c complex, the MMS-X interactive graphics (Barry *et al.*, 1976) with the macromolecular modelling system M3, developed by C. Broughton (Sielecki *et al.*, 1982), was used for map interpretation and model fitting.

The initial electron density map from the molecular replacement solution was of high quality and allowed the subtilisin Carlsberg molecule to be recognized and ~~fitted~~ easily to the density. The electron density for the majority of the amino acid side-chains differing between subtilisin Novo and subtilisin Carlsberg could be interpreted correctly, indicating that model bias was not affecting the map strongly. Two regions had weak electron density: the exterior loop in which Thr55 is deleted from Novo to Carlsberg (Fig. 2.1), and the exterior loop Gly157-Thr162. The entire subtilisin Carlsberg molecule was fitted to the electron density map before refinement of the complex was begun.

This map had relatively poor electron density in the region that was assigned to eglin-c on the basis of packing considerations. The structure of CI-2 (McPhalen *et al.*, 1985) was used as a guide in fitting the inhibitor; with this guide, residues 43I to 71I and 75I to 83I of eglin-c were fitted through some discontinuities in the density. The weakest regions of inhibitor density were in loops at the edge of the molecule, as was the case with CI-2. An ideal α-helix of poly-L-alanine was fitted to a discontinuous but clearly α-helical segment of density in the region of Thr31I to His42I. These 50 residues were added to the molecular model with the refitted subtilisin Carlsberg and a second electron density map was calculated with the augmented model. The remainder of the eglin-c molecule, Lys22I to Lys30I and Pro72I to Asn74I, was fitted to the second map and the poly-L-alanine helix was replaced with the correct sequence of amino acids from the inhibitor. These models of subtilisin Carlsberg and eglin-c were used to begin least-squares refinement of the complex. One strong peak was observed in the initial electron density map in a position close to that of Ca^{2+} ion 1 in the subtilisin Novo:CI-2 complex, although the Ca^{2+} ion had not been included in the phasing model. On the assumption that the peak corresponded to an ion attached to the subtilisin Carlsberg molecule, it was included in the refinement of the complex as a sulphur atom, with no restraints on its non-bonded contact distances.

The restrained-parameter least-squares refinement program of Hendrickson and Konnert (1980), modified locally by M. Fujinaga for the FPS164 attached processor, was used to refine the atomic parameters of the subtilisin Carlsberg:eglin-c complex. The parameters that were restrained and the general strategy used for the refinement are the same as those used for the subtilisin Novo:CI-2 complex (section 2.1.4). The course of this refinement is summarized in Table 3.2 and the changes in the *R*-factor during the refinement are shown in Figure 3.1. Data from crystal 1 were used for the first 22 cycles, and data from crystal 2 for cycles 23 to 57 (Table 3.1). As with the subtilisin Novo:CI-2 complex, the subtilisin Carlsberg molecule was similar enough to the molecular replacement model that large manual adjustments during refitting were unnecessary, even in early stages of the refinement. The eglin-c molecule was adjusted more extensively, particularly in the exterior loops with weak density, but its folding is basically the same as that of the CI-2 model used for the

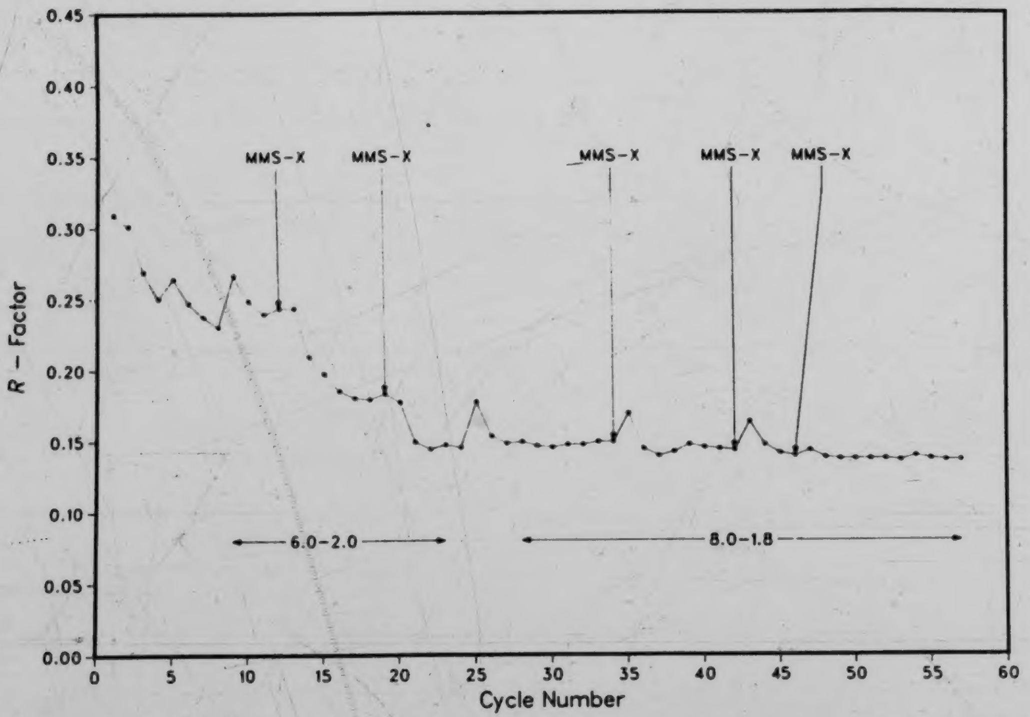


Figure 3.1. Progress of the Subtilisin Carlsberg:Eglin-c Complex Refinement. The *R*-factor at each cycle is plotted versus cycle number. The resolution range of the data included at each cycle is indicated. Data from 6.0 to 2.8Å were used for cycles 1 to 4, 6.0 to 2.5Å for cycles 5 to 8, 8.0 to 2.0Å for cycles 23 to 24, and 8.0 to 1.9Å for cycles 25 to 27. Electron density maps with coefficients $2m|F_o|-D|F_c|$, α_c were calculated and used to refit the model on the graphics system at the points marked "MMS-X".

initial fitting. At cycle 42, the sidechains of Tyr104, His41I, Tyr63I and Asn74I remained poorly determined. For cycles 43 to 46 these four residues were refined as alanines. Ideal sidechains were fitted to an electron density map calculated after cycle 46; this map showed improved density in these regions.

Solvent molecules were chosen from difference electron density maps and added to the model beginning at cycle 20. At cycle 35, two water molecules with occupancies of 1.0 and

Table 3.2

Course of Least-squares Refinement

Cycle number	Data used(Å)	Number of reflections	R-factor	Comments
1	6.0-2.8	6498	0.309	Start refinement with model containing entire subtilisin Carlsberg sequence', residues 22I to 83I of inhibitor, 1 ion in position of Ca^{2+} ion 1 of subtilisin Novo. Ion refined as sulphur atom. Use all reflections with $I \geq \sigma(I)$. One overall B-factor.
5	6.0-2.5	9321	0.264	Add in data from 2.8 to 2.5Å.
9	6.0-2.0	17825	0.266	Add in data from 2.5 to 2.0Å.
12	6.0-2.0	17825	0.243	Refit enzyme and inhibitor.
13	6.0-2.0	17825	0.243	Restart refinement with individual atomic B-factors.
19	6.0-2.0	17825	0.183	Refit enzyme and inhibitor. Calculate difference map, coefficients $m\lvert\mathbf{F}_O\rvert-\lvert\mathbf{F}_C\rvert$, α_C. Choose solvent positions from difference map peaks.
20	6.0-2.0	17825	0.177	Restart refinement with 90 H_2O and 1 ion in model.
23	8.0-2.0	18265	0.147	Add data from 8.0 to 6.0Å.
25	8.0-1.9	23291	0.177	Add data from 2.0 to 1.9Å.
28	8.0-1.8	27094	0.149	Add data from 1.9 to 1.8Å.
34	8.0-1.8	27094	0.149	Refit enzyme and inhibitor. Remove H_2O with high B-factors and low occupancies. Choose new solvent from difference map. Add Leu21I to N-terminus of inhibitor.
35	8.0-1.8	27094	0.169	Restart refinement. Include 2 more well-defined solvent molecules as ions, refine as sulphur atoms. Total of 138 H_2O and 3 ions in model.
42	8.0-1.8	27094	0.143	Refit enzyme and inhibitor. Remove H_2O with high B-factors and low occupancies. Choose new solvent from difference map. Poor side-chain density for residues Tyr104, His42I, Tyr63I, and Asn74I; replace with alanine residues.
43	8.0-1.8	27094	0.163	Restart refinement with 165 H_2O and 3 ions in model.

(Table 3.2 continued)

46	8.0-1.8	27094	0.139	Add side-chains replaced with alanine at cycle 42 back to model, fitting to new electron density map. Add 1 H_2O to model, close to Tyr104.
47	8.0-1.8	27094	0.143	Restart refinement. Refine 3 ions as Ca^{2+}.
57	8.0-1.8	27094	0.136	End of refinement.

---

[1] Smith *et al.*, 1968.

unusually low *B*-factors were converted into ions, to give a total of 3 ions in the model. All 3 ions were refined as sulphur atoms to cycle 46; based on their refined occupancies and *B*-factors and their coordination geometry, they were subsequently refined as Ca^{2+} ions. The water molecules from the final refinement cycle were ordered according to the empirical 'quality factor' of James & Sielecki (1983), defined as occupancy^2/*B*. Thus, water molecules with low sequence numbers are relatively more reliable than those with high sequence numbers. The lowest sequence number for a water molecule in this complex is 342.

3.1.5 Quality of the Refined Complex Structure

The refinement of the subtilisin Carlsberg:eglin-c structure was considered complete after 57 cycles. The model had essentially stopped changing and there were few indications that further changes in the model were necessary. This section describes the criteria used to judge the completeness of the refinement and the quality of the final structure.

The final refinement parameters are given in Table 3.3. The restraints on this final structure are tight, and the indicated parameter shifts for this final cycle are very small. The rms coordinate shift is 0.005Å; the maximum coordinate shifts are 0.044Å for atoms in the enzyme, 0.033Å for atoms in the inhibitor, and 0.040Å for solvent molecules. The vast majority of the coordinate shifts for both

Table 3.3

Final Refinement Parameters and Results

No. of cycles	57
R-factor	0.136
Resolution range (Å)	8.0-1.8
No. of reflections ($I \geq \sigma(I)$)	27094
No. of protein atoms	2450
No. of solvent atoms[1]	167
No. of variable parameters	10636
$\langle\|\|F_o\|-\|F_c\|\|\rangle$	17
<Coordinate shift> (Å)	0.005
<*B*-factor shift> (Å^2)	0.22
Rms deviations from ideal values[2]	
Distance restraints (Å)	
Bond distance	0.009(0.008)
Angle distance	0.027(0.016)
Planar 1-4 distance	0.031(0.016)
Plane restraint(Å)	0.017(0.012)
Chiral-center restraint (Å^3)	0.146(0.080)
Non-bonded contact restraints (Å)	
Single torsion contact	0.268(0.400)
Multiple torsion contact	0.176(0.400)
Possible hydrogen bond	0.147(0.400)
Isotropic thermal factor restraints (Å^2)	
Main-chain bond	1.954(2.000)
Main-chain angle	2.506(2.000)
Side-chain bond	3.897(3.000)
Side-chain angle	5.329(3.000)

[1] Including 3 Ca^{2+} ions.
[2] The values of σ, in parentheses, are the input estimated standard deviations that determine the relative weights of the corresponding restraints [see Hendrickson and Konnert (1980)].

protein and solvent atoms are less than 0.010Å. The restraints on this final structure are the same as the ones on the final refined structure of the subtilisin Novo:CI-2 complex. The indicated parameter shifts for the subtilisin Carlsberg:eglin-c complex are very close to those of the Novo:CI-2 complex; the maximum coordinate shifts are approximately half those of the Novo:CI-2 complex. In general, the refinement of the subtilisin Carlsberg:eglin-c complex proceeded more smoothly and rapidly to convergence. This is probably due to the use of an almost fully refined model in the molecular replacement, one with few conformational differences from the new structure, and one that accounts for a high percentage of the expected scattering matter in the new structure.

A difference electron density map was computed at cycle 57 and examined on the graphics system. The highest peak on the difference map was 0.22e/Å^3.[2] The map contained 7 additional peaks with heights greater than 0.20e/Å^3, and 35 negative troughs with depths greater than 0.20e/Å^3. The eight positive peaks are most likely additional solvent molecules. Five of the negative troughs are associated with side-chains of amino acids on exterior loops with high temperature factors; seven troughs lie close to main-chain atoms in these regions. In such regions, negative troughs may indicate underestimation of temperature factors due to the relatively weak electron density. Three small troughs are associated with one of the calcium-binding loops in subtilisin Carlsberg, Leu75 to Val80. The remainder of the negative troughs are associated with segments of the model that are not properly centered in the electron density; the density is distorted and does not conform to a model with good geometry. The majority of these deviations are small.

---

[2] All electron density maps were calculated omitting the contribution of the F_{000} term and thus have a mean electron density of zero. Contour levels therefore do not refer to the actual electron density but rather to the deviation above the mean density of the map. Contour levels for difference maps do refer to actual electron density because the F_{000} term is common to calculated and observed structure factors.

After the refinement of the subtilisin Carlsberg:eglin-c complex was completed, the DNA sequence of a subtilisin Carlsberg-like alkaline proteinase from *B. licheniformis* was published (Jacobs *et al.*, 1985). The culture medium of this bacterium was tested for the presence of the enzyme with antibodies raised against commercial subtilisin Carlsberg. Keay & Moser (1969) have shown, on the basis of kinetic properties and immunological cross-reactivity, that several species of *Bacillus* produce enzymes that belong to the group of Carlsberg-like enzymes. Antibodies raised against this family do not cross-react with subtilisin Novo-like enzymes. The DNA sequence of the *B. licheniformis* enzyme is different from the published protein sequence (Smith *et al.*, 1968) in 5 places. Careful examination of the electron density and difference maps from cycle 57 showed that in one case the DNA sequence was probably correct; the side-chain of residue 158 has density much more appropriate for a serine than an asparagine. In the other 4 cases, Ser103, Ala129, Asn212, and Ser161, there are no indications in either map that the residue assigned by the protein sequence is incorrect. The information from the X-ray structure must be treated with caution, however, because the side-chains of these 4 residues are somewhat less ordered than average. It is possible that the protein used in these structural studies has a slightly different sequence than that from the organism providing the DNA sequence. The protein used in our studies may be produced by a different *Bacillus* or a different strain of *B. licheniformis*. Since our protein is a commercial product, its origins presumably are a closely guarded secret.

An estimate of the accuracy of the coordinates from a

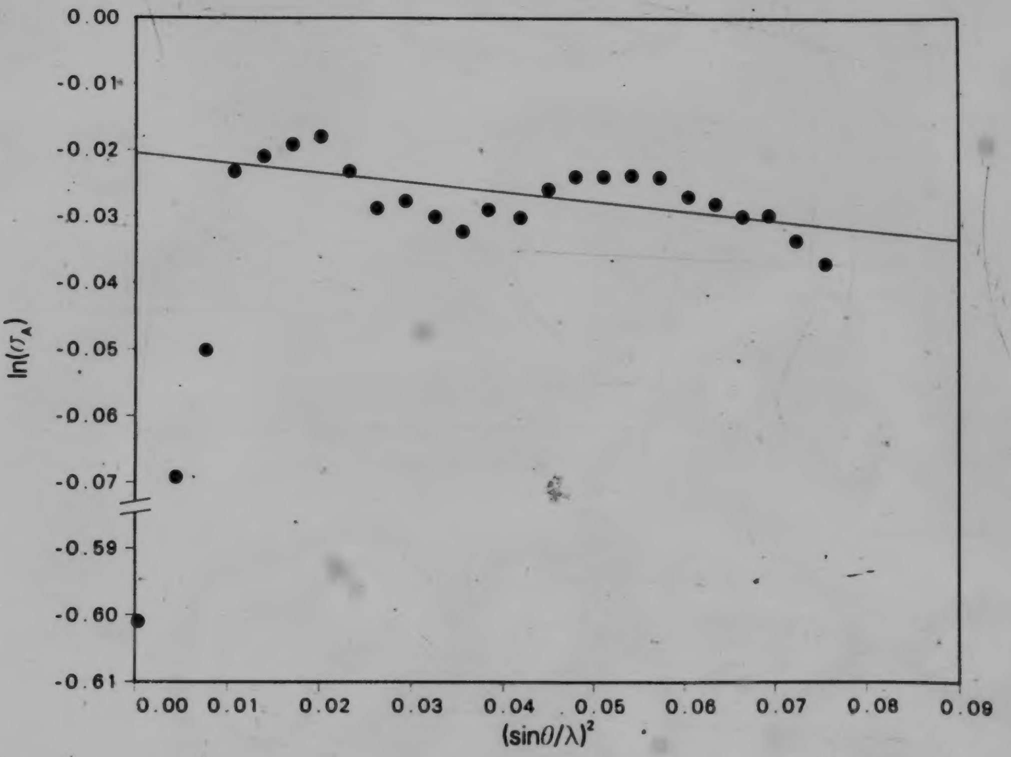


Figure 3.2. σ_A Plot to Estimate Coordinate Error. The slope of the straight-line portion of this plot is $[-\pi^3(<|\Delta r|>)^2]$ (mean coordinate error), or $[-(8\pi^2/3)<|\Delta r|^2>]$ (rms coordinate error). The line is determined by a least-squares fit to all data points but the first three. The first three data points describe low resolution data that is affected by omission of bulk solvent from the protein model.

crystal structure is important to the assessment of the quality of the structure. The accuracy of the atomic coordinates in the fully refined subtilisin Carlsberg:eglin-c structure was estimated with two different methods: the σ_A plot of Read (1986) and the method of Cruickshank (1949, 1954, 1967). The σ_A plot gives an overall estimate of the coordinate errors in the structure based on a derivation similar to that of Luzzati (1952), but with a somewhat different set of basic assumptions (section 2.1.5). The σ_A plot for the subtilisin Carlsberg:eglin-c complex is given in Figure 3.2. The overall mean coordinate error for the structure from this method is 0.07Å, corresponding to an rms error of 0.08Å. These estimated errors are somewhat smaller than those observed for other highly refined protein structures; for structures of a similar size, resolution range, and *R*-factor, the errors are in the range of 0.10 to 0.15Å (James & Sielecki, 1983; Read *et al.*, 1983). The estimated errors for the subtilisin Carlsberg:eglin-c complex are about half those for the subtilisin Novo:CI-2 complex, but at least part of this difference is due to the higher resolution of the data for the subtilisin Carlsberg:eglin-c complex.

The formulae of Cruickshank (1949, 1954, 1967) may be used to obtain an estimate of individual atomic coordinate errors in the structure. The accurate use of these equations requires, however, that the axes of the cell be close to orthogonal so that the coordinate errors along each

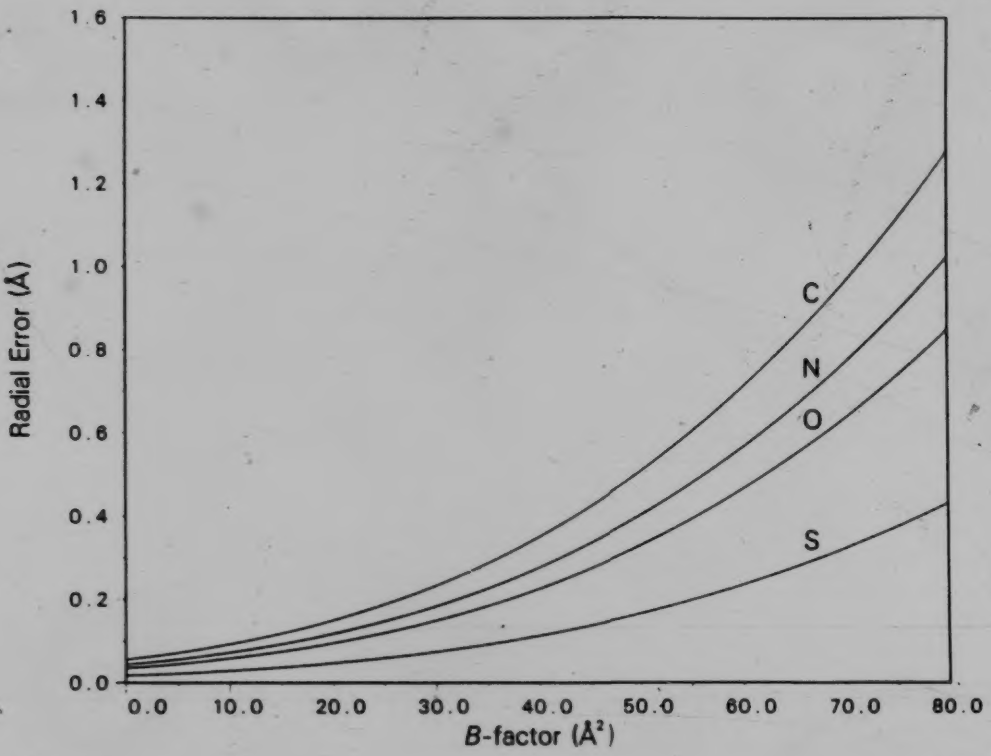


Figure 3.3. Atomic Coordinate Error by the Method of Cruickshank. The estimated radial standard deviations in atomic position are given as a function of *B*-factor. The 4 curves, from top to bottom, are for carbon, nitrogen, oxygen, and sulfur atoms in the final refined structure of subtilisin Carlsberg and eglin-c. No curve is shown for Ca^{2+}; the 3 Ca^{2+} ions have estimated radial standard deviations of 0.03Å.

axis are not intercorrelated. This assumption does not hold for the crystals of this structure, with cos α = 0.351, cos β = 0.111 and cos γ = 0.255. With this probable source of error in mind, the coordinate errors along each axis were calculated with the Cruickshank formula for a non-centered monoclinic space group (section 2.1.5). Radial errors were calculated for each atom type in the structure and for a range of *B*-factors, with $\sigma(r_i) = \sqrt{3}\sigma(y_i)$. A plot of radial error as a function of *B*-factor is given in Figure 3.3 for the carbon, nitrogen, oxygen, and sulphur atoms in the subtilisin Carlsberg:eglin-c complex. The data in this plot may be used in conjunction with Figure 3.4, showing *B*-factors as a function of position along the polypeptide chain, to obtain an estimate of the accuracy of a specific region of the structure. Regions with main-chain temperature factors above the mean for the complex are all external loops. Residues forming the active site of the enzyme or the reactive site loop of the inhibitor have some of the lowest main-chain temperature factors in the complex.

An overall estimated coordinate error can be determined from the method of Cruickshank by calculating the rms value of the estimated radial errors for all the atoms in the model. This value for the subtilisin Carlsberg:eglin-c complex is 0.12Å, close to the value of 0.08Å obtained from the method of Read.

A second measure of the quality of a crystal structure is the deviation of the geometry of the model structure from ideal values. Ideal values of geometric parameters may be obtained from small molecule crystal structures (Sielecki *et al.*, 1979). Values for deviation from ideality of some geometrical parameters for this complex are given in Table 3.3; a more detailed summary of the peptide bond geometry for the subtilisin Carlsberg and eglin-c molecules is given in Table 3.4. The deviations from ideal values for the

Table 3.4

Summary of Peptide Bond Geometry

	Bond type	No.	Mean	RMS deviation	Minimum deviation	Maximum deviation	Expected[1] value
Subtilisin Carlsberg	N-C$^{\alpha}$(Å)	274	1.476	0.009	1.448	1.508	1.474
	C$^{\alpha}$-C(Å)	274	1.526	0.008	1.501	1.548	1.524
	C=O(Å)	274	1.244	0.007	1.226	1.264	1.243
	C-N(Å)	273	1.325	0.008	1.307	1.389	1.323
	τ(°)	274	110.4	3.0	103.1	122.3	109.65
	ω(°)	271[2]	-179.9	3.4	-165.1	171.0	180.00
Eglin-c	N-C$^{\alpha}$(Å)	64	1.474	0.008	1.450	1.493	1.474
	C$^{\alpha}$-C(Å)	64	1.526	0.008	1.512	1.550	1.524
	C=O(Å)	64	1.244	0.006	1.232	1.257	1.243
	C-N(Å)	63	1.322	0.006	1.309	1.337	1.323
	τ(°)	64	110.2	3.1	103.8	118.6	109.65
	ω(°)	63	180.0	3.1	-174.3	169.8	180.00

[1] Expected values from Arnott & Dover (1967).

[2] The peptide bonds before Pro168 and Thr211 are *cis*.

subtilisin Carlsberg:eglin-c complex are within the range of deviations found for small molecule structures. The mean values of the observed parameters in Table 3.4 deviate from the expected values by less than 0.4 times the rms deviation. The deviation of minimum or maximum values from the mean exceeds 4 times the rms deviation in only one case, the maximum observed C - N bond length. The agreement of the geometric parameters of this structure with ideal values is as good or better than that of the subtilisin Novo:CI-2 complex; the deviation of minimum and maximum values from the mean is slightly higher for this structure.

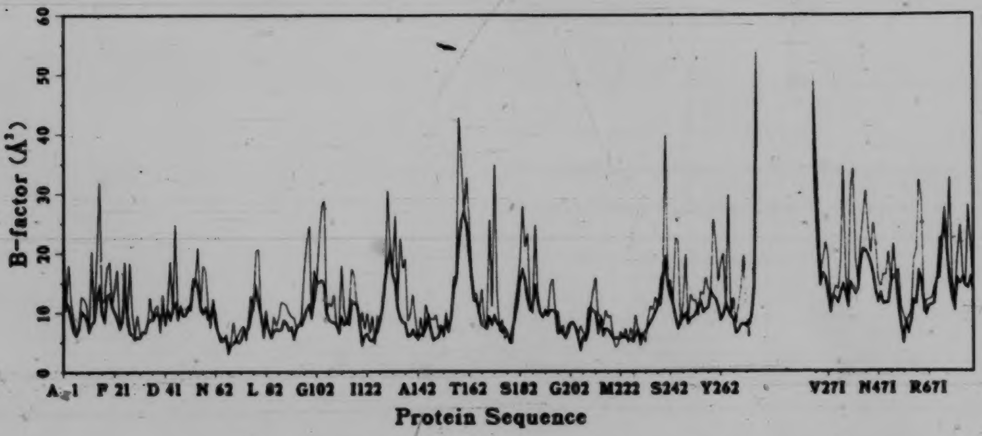


Figure 3.4. Variation in *B*-factor Along the Polypeptide Chain. The heavy line denotes the mean *B*-factor of the main-chain atoms, the light line that of the side-chain atoms. The vertical line separates enzyme and inhibitor residues. Inhibitor residue numbers are followed by an 'I' to distinguish them from those of the enzyme. No *B*-factors are given for the 6 N-terminal residues of the inhibitor; they are not seen in the electron density map. The mean overall *B*-factor for the atoms of the complex is 12Å².

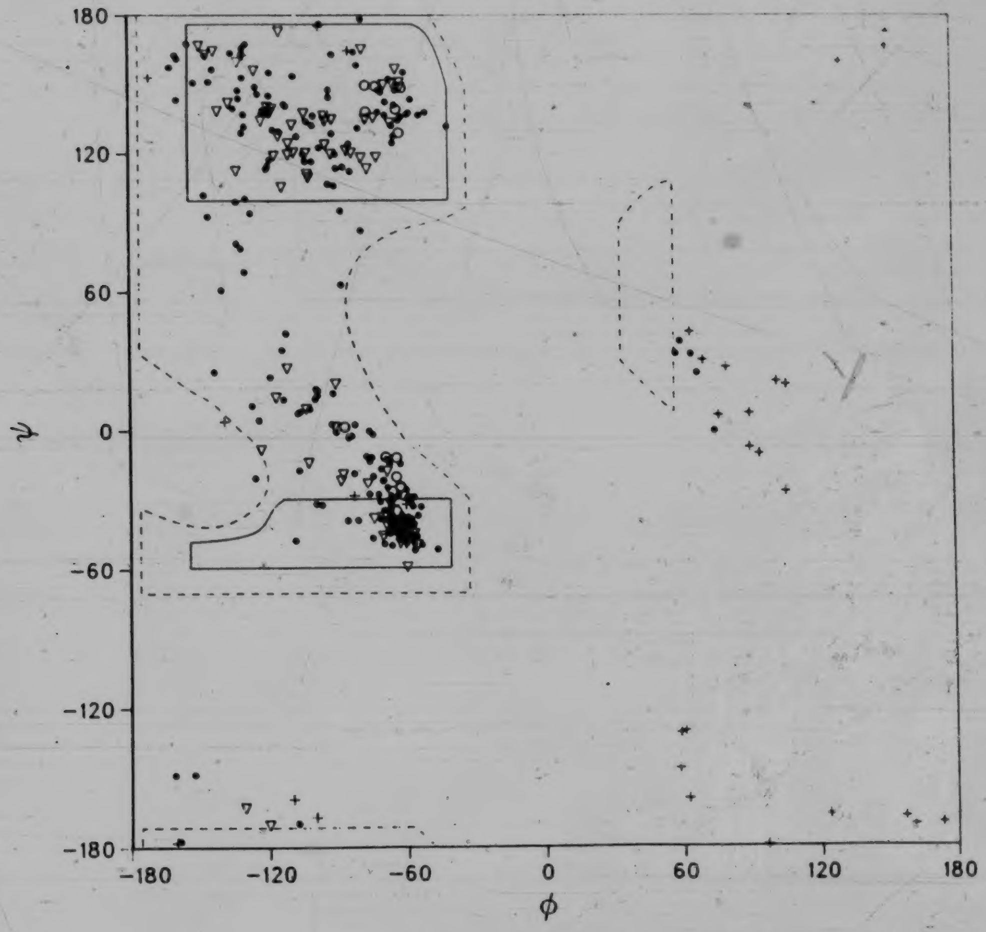


Figure 3.5. ϕ-ψ Plot for Subtilisin Carlsberg and Eglin-c. The solid lines enclose the fully allowed conformational regions for τ (N-C$^{\alpha}$-C) of 110°; the dotted lines enclose the more permissive regions of smaller acceptable van der Waals' contacts for τ of 115° (Ramakrishnan & Ramachandran, 1965). The symbols denote prolines (o), β-branched amino acids (Δ), glycines (+), and all other amino acids (•). Residues outside the allowed conformational regions and their ϕ-ψ angles are Lys12 (64,33), Asn25 (74,0), Asp32 (-161,-149), Asn77 (-153,-149), Val81 (-120,-170), Asp181 (-108,-170), Ser184 (67,25), Asn212 (58,33), Thr213 (-131,-163), Leu257 (-125,-21), and Asn74I (59,38).

The quality of the geometry of a structure may also be tested by examining how well its main-chain torsional angles conform to allowed values, calculated from minimum contact radii of non-bonded atoms (Ramakrishnan & Ramachandran, 1965). The distribution of the main-chain torsion angles for the entire complex is presented in Figure 3.5, a Ramachandran plot. The torsion angles for most residues fall well within the conformationally allowed regions, with major clusters in the α-helix and β-sheet areas. The 11 residues other than glycines that fall outside the allowed regions are mainly found in irregular turns or segments of unclassified structure. Five of these are in segments of chain with higher *B*-factors and their geometry may be more poorly defined. Two of them are Asn77 and Val81, ligands of Ca^{2+} ion 1; these two residues also lie outside the allowed regions in subtilisin Novo. Two others follow Thr211, one of the residues preceded by a *cis*-peptide bond, and are part of a 5-residue turn. One is Asp32, the catalytic aspartic acid, with $\phi = -161°$ and $\psi = -149°$. In the latter 2 cases, the unusual torsion angles may be due to strain imposed by stringent structural requirements, such as the precise positioning of the side-chain of Asp32. Asp32 in subtilisin Novo also falls outside the allowed conformational regions. The single residue of eglin-c outside the allowed regions is Asn74I, and Asp74I has similar torsional angles in CI-2.

The final electron density map calculated for the subtilisin Carlsberg:eglin-c complex at cycle 57 of the least-squares refinement is generally of excellent quality. Some regions of weak density remain; these are described in Table

Table 3.5

Residues with Poor Electron Density at Cycle 57

Residue	Comment
Lys12	Weak density for C^{δ}, N^{ζ}, none for C^{ϵ}.
Lys15	No density past C^{γ}.
Lys27	No density for N^{ζ}.
Val45	Weak density for C^{γ}'s.
Tyr104	Weak density for C^{ζ}, none for $C^{\delta 2}$, $C^{\epsilon 2}$, OH.
Ser105	Weak density for O^{γ}.
Gln112	Weak density between C^{γ} & C^{δ}.
Thr116	Weak density for C^{γ}.
Ser130	Weak density for O^{γ}.
Lys136	Weak density for C^{η} & N^{ζ}.
Asn158	No density beyond C^{γ}. May be serine according to DNA sequence.[1]
Ser159	Weak density for C^{α}, C^{β}, O^{γ}.
Gly160	Weak density for C^{α}.
Ser161	Weak density for O^{γ}.
Thr162	Weak density for C^{δ}.
Lys170	No density beyond C^{δ}.
Asp172	No density for O^{δ}.
Asn240	No density for C^{γ}, $O^{\delta 1}$.
Asn248	Weak density between C^{β} & C^{γ}.
Lys265	No density beyond C^{δ}.
Leu21I	Weak density for all atoms.
Lys22I	No density beyond C^{γ}.
Asp33I	No density for $O^{\delta 2}$.
Arg36I	Weak density for guanidinium group.
Glu37I	No density between C^{γ} & C^{γ}.
Leu41I	No density for $C^{\delta 2}$.
His42I	Weak density for C^{δ} & C^{ϵ}.
Tyr63I	Weak density for atoms of ring.
Asn64I	Weak density for O^{δ}s.
Gly72AI	No density for C^{α}.
Asn74I	Weak density for O^{δ}s.
His81I	Weak density for guanidinium group.

[1] Jacobs *et al.*, 1985.

3.5. As with the subtilisin Novo:CI-2 complex, all of the residues with weak density are on the surface of the enzyme or the inhibitor, and many have side-chains protruding out into the solvent. All residues were examined for weak density at a contour level of 0.56e/Å^3 to facilitate comparison with the subtilisin Novo:CI-2 complex. Examples of regions of good and poor density are given in Figure 3.6. The electron density for six residues in the active site, a region of good density, is shown in Figure 3.6a. The poorly ordered external loop, Asn158 to Thr162, a region of very weak density, is shown in Figure 3.6b.

One final indication of the quality of a crystallographic structure is the *R*-factor. The *R*-factor for the fully refined structure of the complex between subtilisin Carlsberg and eglin-c, for all data in the resolution range from 8.0 to 1.8Å with $I \geq \sigma(I)$, is 0.136. This value was calculated with 27094 reflections, 93% of the unique set of reflections measured. The *R*-factor calculated on the complete unique set is 0.164. The criteria of quality discussed in this section indicate that the structure of this complex has been refined to convergence, that the errors in the atomic parameters are small, that the geometric parameters do not deviate greatly from ideal values, and that the agreement between observed and calculated structure factors is good.

3.2 The Structure of Subtilisin Carlsberg

3.2.1 Secondary and Tertiary Structures

The folding pattern of the backbone of the subtilisin Carlsberg molecule is almost identical to that of subtilisin

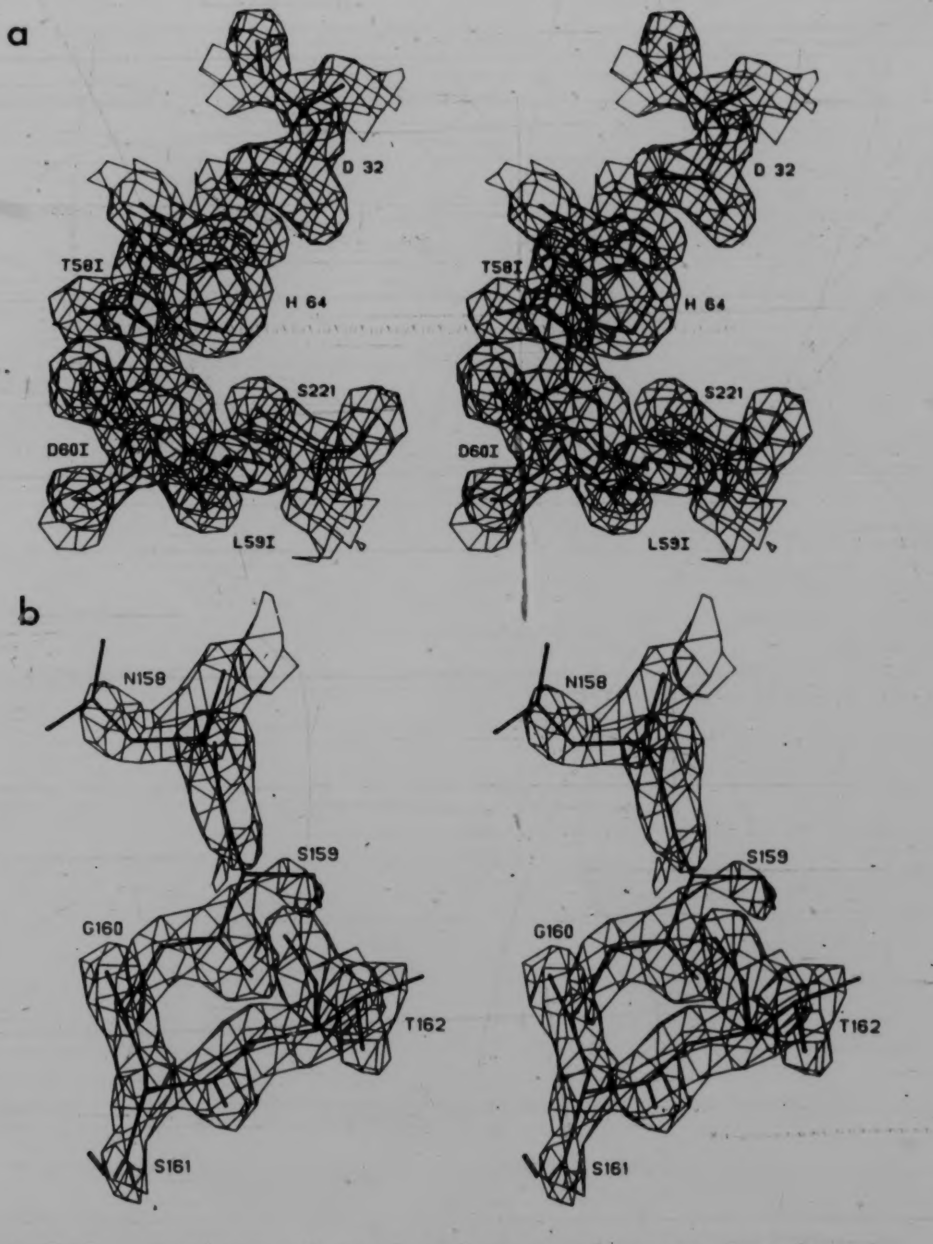
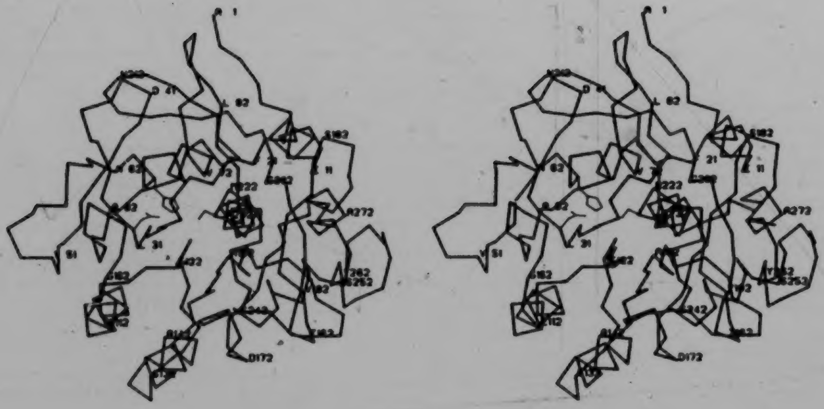
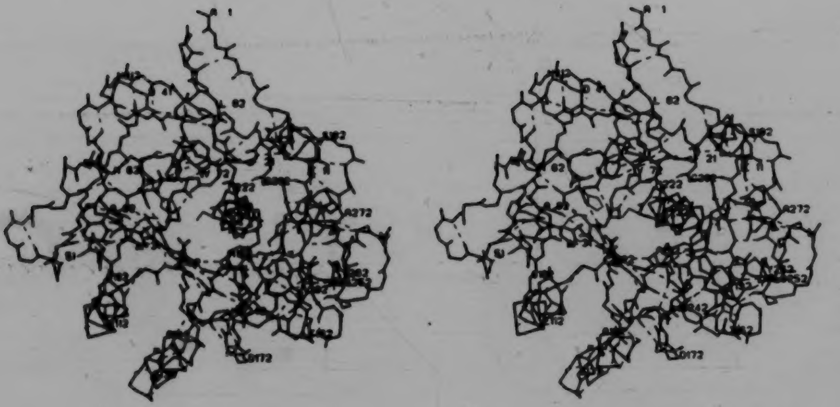

a
D 32
T58I
H 64
S221
D60I
L59I
b
N158
S159
G160
T162
S161

Figure 3.6. Regions of Good and Poor Electron Density. Both figures are contoured at a level of $0.56e/\text{Å}^3$. a) Residues Asp32, His64, and Ser221 from the subtilisin Carlsberg active site, and Thr58I to Asp60I from the reactive site loop of the inhibitor. The active site bond lies between Leu59I and Asp60I. b) The external loop Asn158 to Thr162 in subtilisin Carlsberg. This region contains the weakest density in the final electron density map and all 5 residues have high *B*-factors.

a


D 41
R272
D172

b



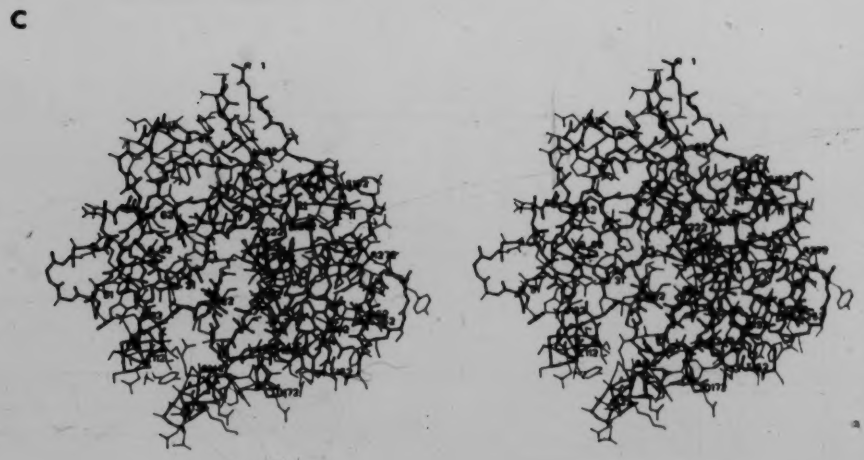


Figure 3.7. Views of Subtilisin Carlsberg. a) An α-carbon backbone drawing of the refined subtilisin Carlsberg molecule from the complex with eglin-c. Side-chains are shown for the catalytic triad, Asp32, His64, and Ser221. The active site is at the front of the molecule. b) The molecule in the same orientation as view (a), showing all main-chain atoms and the hydrogen bonds between them (dashed lines). c) The molecule in the same orientation as view (a), with side-chains added for all residues. Thick lines denote the main-chain, thin lines the side-chains. Every 10th α-carbon is labelled in each view.

Novo as described in chapter 2. The central 7-stranded parallel β-sheet is conserved, as well as the α-helices packed against the sheet and running antiparallel to it (Fig. 3.7). The general topology of subtilisin Carlsberg is the same as that depicted for subtilisin Novo in Figure 2.13. The hydrogen-bonding patterns of the two molecules are not identical, however, so the extent of some of the secondary structural elements in subtilisin Carlsberg is slightly different (Table 3.6). The same methods and criteria were used in the assignment of hydrogen bonds and secondary structural elements in subtilisin Carlsberg as in subtilisin Novo.

The strong conservation of topology between the two enzymes is not entirely surprising; most of the α-helices and the strands of β-sheet comprise the molecular core and this is the region of the structures with very few amino acid changes. With knowledge of the structure of the native subtilisin Novo and the positions of the sequence changes between the two enzymes, Wright *et al*. (1969) predicted that the three-dimensional structure of subtilisin Carlsberg would strongly resemble that of subtilisin Novo. The strong conservation of turns and turn types on the surfaces of the molecules is more surprising, since these regions would be expected to be more conformationally free and labile, and most of the sequence changes are for surface residues.

The hydrogen bonds in the subtilisin Carlsberg:eglin-c complex have been grouped according to the types of atoms forming them and the type of secondary structure to which they belong. The mean geometric parameters of these groups are summarized in Table 3.7. As in the subtilisin Novo molecule, the mean values of the parameters are the same for the different types of secondary structure, within the limits of accuracy of the atomic coordinates. The mean values are also the same for intra-main-chain bonds versus those involving side-chain or solvent molecules. The mean values of all the geometric parameters listed in Table 3.7 are statistically the same for subtilisin Novo and

Table 3.6

Secondary Structural Elements of Subtilisin Carlsberg

Element	Residues Involved
Parallel β-sheet	(1) Val26-Asp32, (2) Gly46-Phe50, (3) Ser89-Val95, (4) Asp120-Met124, (5) Val148-Ala153, (6) Ile175-Val180, (7) Val198-Gly202.
Antiparallel β-sheet	Val205-Tyr209, Thr213-Leu217.
β-bridges[1]	Gly128, Tyr167, Ala179-Asp181, Asn185-Ala187, Thr255, Gly266-Ile268.
α-helices	(A) Tyr6-Ile11, (B) Ala13-Ala18, (C) Gly63-Ala74, (D) Ser103-Asn117, (E) Ser132-Arg145, (F) Ser224-His238, (G) Ser242-Ser252, (H) Ser259-Gly264, (I) Asn269-Ala274.
Helical turn	Thr220-Ser224.
Type I turns[2]	Pro9-Lys12, Gln36-His39, His39-Leu42, Ala85-Val88, Asn97-Gly100, Tyr143-Gly146, Tyr171-Val174, Asp181-Ser184, Ala187-Ser190, Gly193-Leu196, His238-Leu241, Ala272-Gln275.
Type II turns	Gly23-Val26, Val51-Glu54.
Type II' turns	Ser159-Thr162, Tyr263-Gly266.
Type III turns	Pro5-Ile8, Val16-Gln19, Gln17-Gly20, Pro168-Tyr171, Gly219-Met222, Thr220-Ala223, Ala223-His226, Ser224-Val227, Ser259-Tyr262.

[1] β structure involving only 2 hydrogen bonds (Kabsch & Sander, 1983).

[2] The turn type is assigned on the basis of the ϕ_2-ψ_2 and ϕ_3-ψ_3 angles (Crawford *et al.*, 1973).

subtilisin Carlsberg.

Because the two subtilisins are so similar in their folding and secondary structural elements, the description of the subtilisin Novo molecule given in section 2.2.1 generally is applicable also to subtilisin Carlsberg (with the appropriate changes in amino acid sequence). One difference between the two subtilisins is the presence of a third ion binding site in subtilisin Carlsberg. This third ion, which may be Ca^{2+} or K^+ (see next section), is bound on

Table 3.7

Hydrogen-bonding Parameters for Subtilisin Carlsberg

Type	No.[1]	N···O(Å) (O···O)[2]	H···O(Å)	N-Ĥ-O(°)
Main-chain to main-chain				
α-Helix(5→1)	57	2.98(0.11)	2.06(0.13)	154(9)
Antiparallel β-sheet	12	2.89(0.09)	1.95(0.11)	157(11)
Parallel β-sheet	33	2.94(0.11)	1.98(0.12)	161(8)
3_{10} turns(4→1)	23	3.04(0.13)	2.16(0.12)	149(13)
Main-chain to side-chain				
N-H···O	31	2.99(0.16)	2.07(0.17)	155(13)
C=O···H-N	18	2.88(0.11)	1.99(0.16)	149(12)
C=O···H-O	24	2.96(0.25)	-[3]	-
Side-chain to side-chain				
N-H···O	15	2.88(0.18)	1.95(0.23)	166(12)
O(H)···(H)O	7	2.75(0.11)	-	-
Main-chain to solvent				
N-H···O	42	2.96(0.12)	2.02(0.14)	158(13)
C=O···O	96	2.87(0.21)	-	-
Side-chain to solvent				
N-H···O	20	2.95(0.20)	2.03(0.21)	156(14)
O(H)···(H)O	65	2.86(0.20)	-	-

[1] Limits for acceptance of a hydrogen bond: donor-acceptor distance ≤ 3.40Å, hydrogen-acceptor distance ≤ 2.40Å, donor-hydrogen-acceptor angle ≥ 125°.

[2] Values in parentheses are the sample standard deviations from the mean.

[3] Values are not given for donors with ambiguous hydrogen atom positions.

the surface of the molecule in the loop between strands 1 and 2 of the central β-sheet. It has 3 carbonyl oxygen ligands from the protein (His39, Leu42, Ala37) and two ordered water ligands. This ion binding site is closely connected with the binding site of ion 1, as Asp41 from the loop defining site 3 is one of the ligands for ion 1.

Subtilisin Carlsberg contains two *cis*-peptide bonds. The first is the bond between Tyr167 and Pro168, that corresponds to a *cis*-peptide found in the same position in subtilisin Novo. The second is between Pro210 and Thr211 (Fig. 3.8); the corresponding residues in subtilisin Novo are Pro210 and Gly211, with a normal *trans*-peptide bond between them. The ϕ-ψ angles for Gly211 in subtilisin Novo are 70° and 40°, not an unusual conformation for a glycine. Both of the *cis*-peptides in subtilisin Carlsberg, as well as the one in subtilisin Novo and the bond between residues

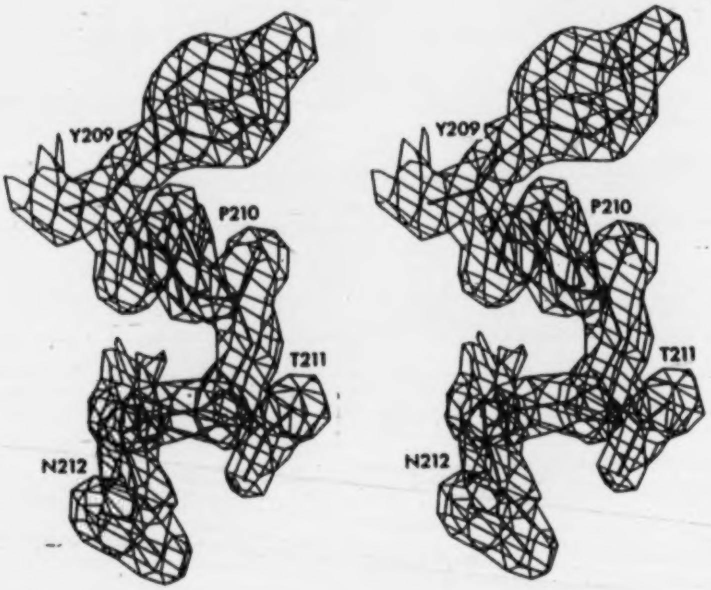
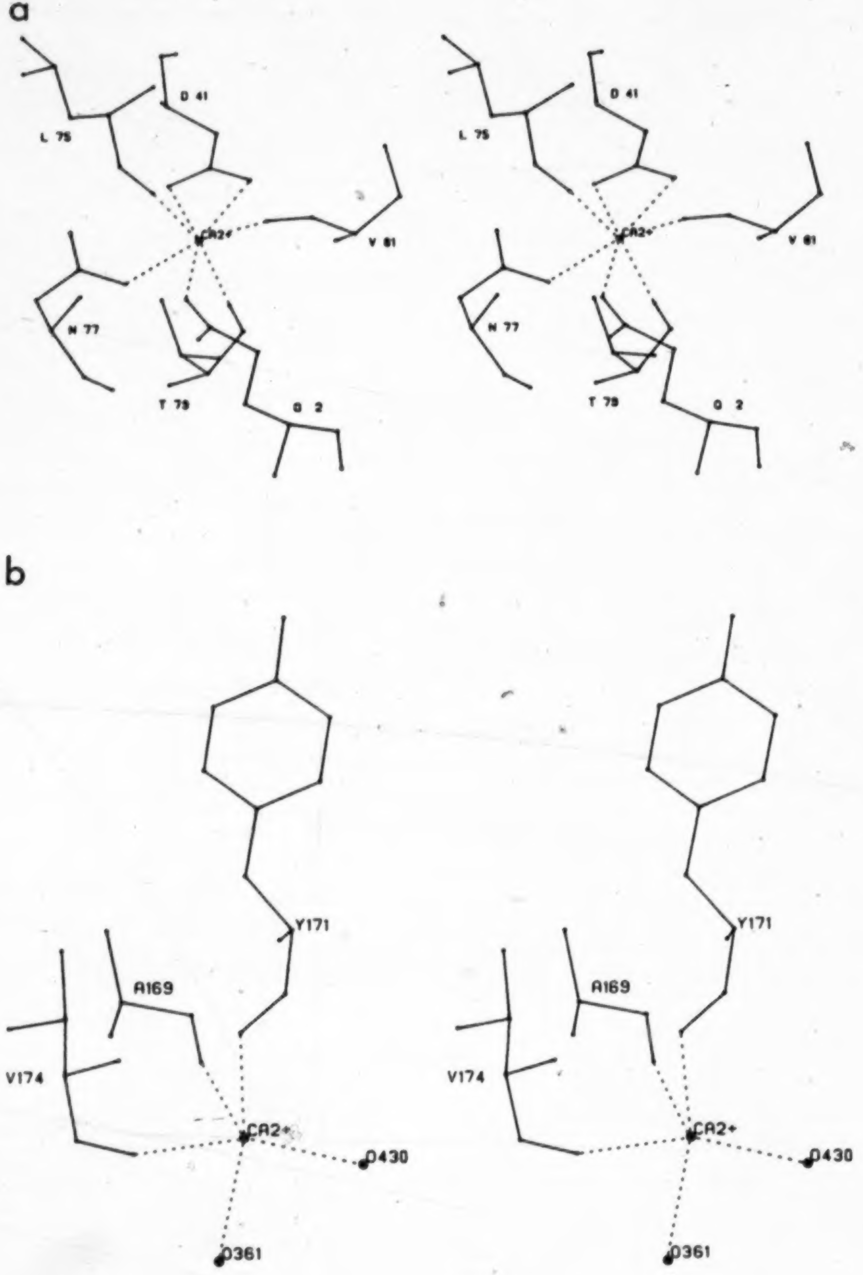


Figure 3.8. The *cis*-Peptide Bond Preceding Thr211. Stereo view of the electron density in the region of the second *cis*-peptide bond in subtilisin Carlsberg, between Pro210 and Thr211. The electron density was contoured at a level of 0.60e/Å³.

Pro210 and Gly211 in subtilisin Novo, are located between residues i+1 and i+2 of 5-membered reverse turns on the surface of the molecule. *Cis*-peptide bonds before proline residues have been observed for about 25% of all proline residues in protein crystal structures; they are found much more rarely before other residues due to energetic and conformational restrictions (Ramachandran & Mitra, 1976). Only 4 *cis*-peptide bonds before residues other than proline have been reported previously in reliable protein crystal structures (Creighton, 1983), although more are being seen as investigators come to believe that they are a rare but reasonable conformation.

3.2.2 Ion Binding Sites

In the initial electron density map calculated from the molecular replacement model for the subtilisin Carlsberg:eglin-c complex, the highest peak in the map was located in the position of Ca^{2+} ion 1 of subtilisin Novo, although this ion had not been included in the phasing model. During the initial stages of refinement, this ion was modelled as a sulphur atom, as was done with the ions bound to subtilisin Novo. Among the solvent molecules chosen from difference maps during the refinement were two more that refined to occupancies of 1.0 and very low *B*-factors as water molecules; these also were refined as ions subsequently. No restraints were applied to non-bonded contact distances for any of the ions during refinement.


a
D 41
L 75
CA2+
V 81
N 77
T 79
Q 2
D 41
L 75
CA2+
V 81
N 77
T 79
Q 2
b
Y171
A169
V174
CA2+
O430
O361
Y171
A169
V174
CA2+
O430
O361

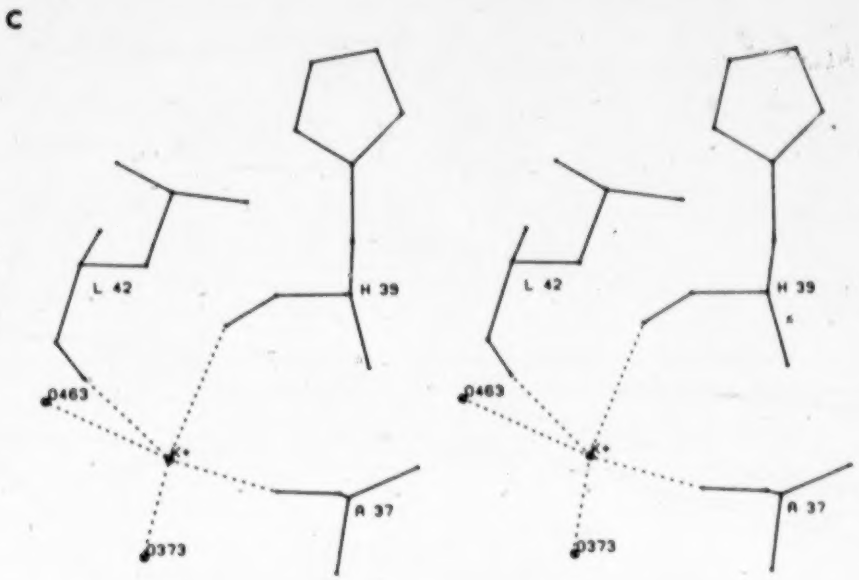


Figure 3.9. Ion Binding Sites in Subtilisin Carlsberg.
a) Residues of subtilisin Carlsberg coordinating Ca^{2+} ion 1. Coordination bonds are indicated with dashed lines.
b) Residues of subtilisin Carlsberg coordinating Ca^{2+} ion 2.
c) Residues of subtilisin Carlsberg coordinating K^+ ion 3. The water ligands are indicated with solid circles.

The coordination geometry of the ions and the nature of their ligands indicated that they were probably either Ca^{2+} or K^+. For the last 11 cycles of refinement, all three ions were refined as Ca^{2+}. The final occupancies and *B*-factors of the 3 ions are: 1.00 and 8Å^2 for ion 1, 0.91 and 10Å^2 for ion 2, and 0.88 and 13Å^2 for ion 3. The coordination geometry of the ions is described in Table 3.8, and each of the ions with their ligands is shown in Figure 3.9.

The average distances for the coordinating ligands to the ions are: 2.37Å for ion 1, 2.56Å for ion 2, and 2.85Å for ion 3. Of the ions tabulated by Brown & Shannon (1973) in their study of metal - oxygen bond lengths and bond

Table 3.8

Geometry of Ion Binding Sites

Coordinating atoms	Ion - O bond length (Å)	Ion - O - C bond angle (°)
Ca^{2+} #1		
Gln2 $O^{\epsilon 1}$	2.39	140
Asp41 $O^{\delta 1}$	2.35	98
Asp41 $O^{\delta 2}$	2.55	89
Leu75 O	2.23	164
Asn77 $O^{\delta 1}$	2.36	133
Thr79 O	2.41	165
Val81 O	2.28	157
Ca^{2+} #2		
Ala169 O	2.57	161
Tyr171 O	2.57	128
Val174 O	2.54	157
O361 O	2.59	-
O430 O	2.53	-
K^+ #1		
Ala37 O	2.81	156
His39 O	2.64	139
Leu42 O	2.58	165
O373 O	3.17	-
O463 O	3.06	-

strengths, only 3 have typical values similar to those of the ions in this complex; Na^+ at 2.42Å, Ca^{2+} at 2.44Å, and K^+ at 2.82Å. Based on the biochemical data showing the stabilization of subtilisin Novo by Ca^{2+}, and the fact that the crystallization media of both complexes contain K^+, only these two ions were considered when the ion types bound to the enzymes were assigned. Octahedral coordination is common for Ca^{2+} ions bound to carboxyl and carbonyl ligands (Einspahr & Bugg, 1981, 1984).

The mean bond length and octahedral coordination geometry of ion 1 indicate strongly that it is a Ca^{2+} ion; in addition, one of its ligands is negatively charged. Ion 2 has only 5 ligands but is relatively accessible to bulk solvent. Its 5 ordered ligands are located approximately in 5 of the 6 possible octahedral coordination positions, and a weakly occupied solvent molecule could occupy the sixth position. Being accessible to solvent, this ion would not require a negatively charged protein ligand to balance its positive charge(s). These considerations combined with the mean coordination bond length for ion 2 indicate that it is also a Ca^{2+} ion. Ion 3 has its ligands arranged similarly to ion 2 and can also be reached by bulk solvent to fill open coordination positions and supply charge balance. Its coordination bond lengths, however, are more typical of a K^+ ion than a Ca^{2+} and it has been labelled as K^+ in tables and figures.

3.3 The Structure of Eglin-c

As with subtilisin Novo and subtilisin Carlsberg, the overall folding of CI-2 and eglin-c from the two complexes is similar. Both molecules as seen in the crystal structures are wedge-shaped disks of about 28x27x19Å, with their reactive site loops at the narrow end of the wedge. CI-2 has 14 more amino acids at its N-terminus than eglin-c according to the amino acid sequences (Fig. 2.2). In the crystal structures of the complexes, however, both

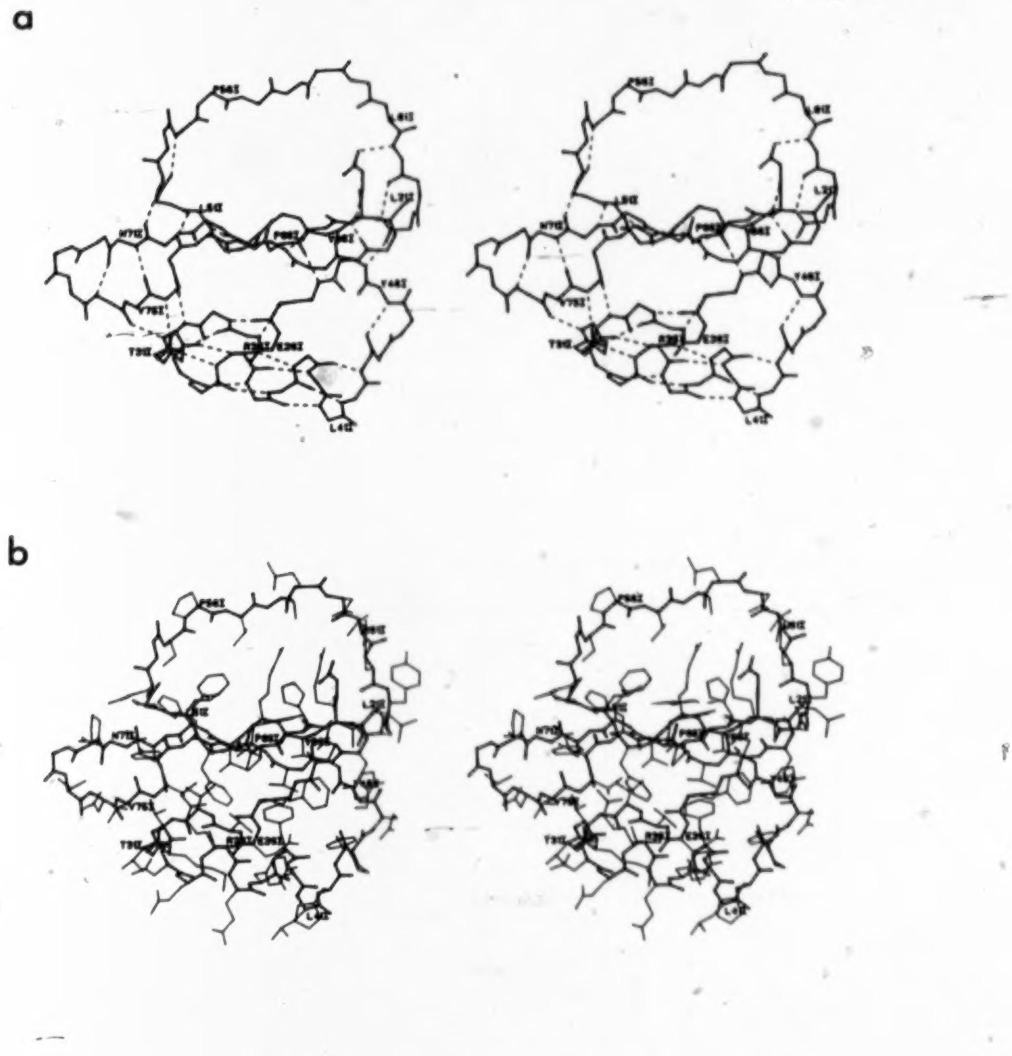


Figure 3.10. Views of Eglin-c. Both views include only residues seen in the electron density map of the subtilisin Carlsberg:eglin-c complex, Leu21I to Gly83I. a) The main-chain atoms of eglin-c and their hydrogen-bonding interactions. Hydrogen bonds are drawn with dashed lines. b) All atoms of the eglin-c molecule. The main-chain is drawn with heavy lines and the side-chains with light lines. Every fifth amino acid residue is labelled in both views.

inhibitors have the same number of observed residues due to a combination of hydrolysis during purification (CI-2) and disorder of the N-termini (both inhibitors). The first residue of eglin-c with visible electron density is Leu21I; the 6 N-terminal residues are not seen. The body of eglin-c is made up of an α-helix of 13 residues, packed against a 'pseudo' 4-stranded mixed parallel and antiparallel β-sheet. Figure 3.11 gives 2 representations of the folding of eglin-c and Table 3.9 lists the secondary structural elements of eglin-c and the hydrogen bonds between the residues comprising those elements.

A comparison of Tables 2.9 and 3.9 shows that the secondary structural elements of CI-2 and eglin-c, the residues comprising those elements, and the geometry of the hydrogen bonds in the elements differ very little. The general description of the molecules is thus similar, and the detailed description of the CI-2 molecule in section 2.3 is also applicable to eglin-c in most respects. The 'pseudo' β-sheet of eglin-c, however, is not identical to the sheet in CI-2. The sheet in eglin-c is composed of Lys22I to Phe24I at one edge with two antiparallel bonds to the C-terminal strand, and strands 1 and 2 of the parallel β-sheet. The C-terminal strand is connected to strand 2 only by a set of well-ordered water molecules participating in hydrogen-bonding bridges linking the two strands. In CI-2, the C-terminal strand is connected to strand 2 by an antiparallel β-bridge in addition to the bridging water molecules (Fig. 2.17).

Table 3.9

Secondary Structural Elements of Eglin-c

Element	Residues[1]	N-H···O length (Å)	N-Ĥ-O angle (°)
Parallel β-sheet	Asn47I-Asn64I	2.77	174
	Val66I-Asn47I	2.94	173
	Tyr49I-Val66I	2.88	159
	Val68I-Tyr49I	2.63	177
	Leu51I-Val68I	2.76	170
	Tyr70I-Leu51I	3.00	149
	Glu53I-Tyr70I	2.80	160
Antiparallel β-bridges	Val82I-Lys22I	2.92	165
	Phe24I-Pro80I	2.88	154
	Val76I-Lys30I	2.95	167
	Val32I-Asn74I	2.92	149
α-helix	Ala35I-Thr31I	2.91	158
	Arg36I-Val32I	2.81	157
	Glu37I-Asp33I	3.11	159
	Tyr38I-Gln34I	3.00	154
	Phe39I-Ala35I	3.04	168
	Thr40I-Arg36I	3.10	160
	Leu41I-Glu37I	2.87	164
	His42I-Tyr38I	2.82	151
	Tyr43I-Phe39I	2.68	146
Type I turns	Val28I-Pro25I	3.10	158
	Tyr46I-Tyr43I	3.10	166
	Arg65I-Arg62I	3.24	153
Type II turns	Lys30I-Val27I	3.06	156
	Ser55I-Pro52I	3.18	149
Type III turn	Val27I-Phe24I	3.12	164
Unclassified	Gly29I-Val76I	2.85	174
	Arg62I-Gly83I	2.85	173
	Asn71I-Val75I	2.99	173
	Asn74I-Asn71I	2.66	130
	Gly83I-Arg65I	2.84	151

[1] The residue listed first in each pair is the hydrogen bond donor.

The hydrophobic core of eglin-c is between the α-helix and the β-sheet. It is composed of residues Phe24I, Val27I,

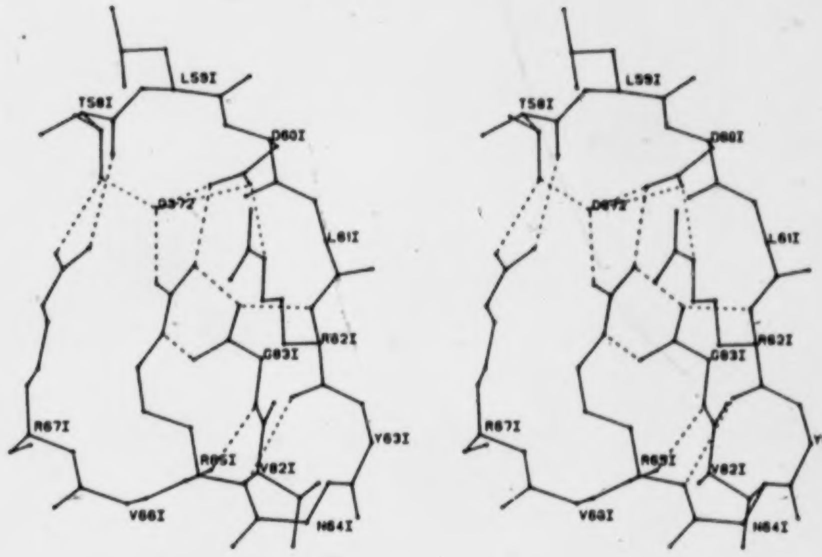

L59I
T58I
D60I
D37I
L61I
R62I
G83I
R67I
Y63I
R65I
V82I
V66I
H64I
L59I
T58I
D60I
D37I
L61I
R62I
G83I
R67I
Y63I
R65I
V82I
V66I
H64I

Figure 3.11. Hydrogen Bonding of the Reactive Site Loop in Eglin-c. Dashed lines indicate hydrogen bonds. Side-chains of residues not involved in hydrogen bonding have been omitted for clarity.

Val32I, Ala35I, Tyr38I, Phe39I, Thr40I, Tyr43I, Tyr46I, Val48I, Phe50I, Val66I, Val68I, Tyr70I, Val76I, Pro80I, and Val82I. The hydrophobic core of eglin-c is formed almost entirely from residues in the same position in the sequence as that of CI-2, but almost 75% of these positions contain amino acid substitutions. The core of eglin-c contains 7 aromatic residues and 10 aliphatics; CI-2 contains 1 aromatic and 15 aliphatics. The single buried polar group in eglin-c is is the C-terminal residue, Gly83I, as in CI-2. The reactive site loop in eglin-c is supported by an extensive network of hydrogen-bonding and electrostatic interactions analogous to those seen in CI-2 (Fig. 3.11). There is one major change between the two reactive site loops, the substitution of Asp60I in eglin-c for Glu60I in CI-2. The differences between the two reactive site regions as a result of the substitution are discussed in chapter 5.

3.4 The Subtilisin Carlsberg:Eglin-c Complex

3.4.1 Inhibitor Binding Interactions

The mode of binding of eglin-c to subtilisin Carlsberg is like that of CI-2 to subtilisin Novo in the relatively small number of residues in the enzymes and inhibitors actually involved in intermolecular contacts (Fig. 2.18). Twelve residues of eglin-c make a total of 134 contacts less than 4.0Å with only 25 residues of subtilisin Carlsberg; these contacts are summarized in Table 3.10.

A similar pattern of enzyme:inhibitor contacts is present in the subtilisin Carlsberg:eglin-c and subtilisin Novo:CI-2 complexes. The majority of contacts are between residues of the reactive site loop and their cognate binding sites on the enzyme. About 60% of the inhibitor contacts involve residues P_4 to P_1, and another 22% involve residues P_1' to P_3'. The active site region of subtilisin Carlsberg and the reactive site loop of eglin-c are shown together in Figure 3.12, which includes most of the residues and contacts given in Table 3.10.

The contacts between eglin-c and subtilisin Carlsberg are more evenly distributed among the residues in the

Table 3.10

Intermolecular Contacts ≤4.0Å for Subtilisin Carlsberg and Eglin-c

	Tyr 49I P_{11}	Leu 51I P_9	Gly 54I P_6	Ser 55I P_5	Pro 56I P_4	Val 57I P_3	Thr 58I P_2	Leu 59I P_1	Asp 60I P_1'	Leu 61I P_2'	Arg 62I P_3'	Arg 67I P_8'	Σ
Thr33							1						1
Asn62											6		6
His64							9	1	4		1		15
Leu96							1						1
Ser99	3											1	4
Gly100						3	5					3	11
Ser101		1			3	1							5
Gly102					4								4
Tyr104			5		1								6
Ile107					2								2
Ser125							1	1					2
Leu126						3		3					6
Gly127				1	3	6		2					12
Gly128			2	2	2								6
Ala129			4										4
Ser130			6										6
Ala152								2					2
Gly154								1					1
Asn155								6	3	1			10
Phe189										3			3
Asn209											3		3
Asn218									3	4			7
Gly219								2					2
Thr220								2					2
Ser221								11	2				13
Σ	3	1	17	3	15	13	17	31	12	8	10	4	134

reactive site loop than in the subtilisin Novo:CI-2 complex. Only 34% of the contacts are made by the P_1 and P_4 residues in the subtilisin Carlsberg:eglin-c complex, compared to almost 50% in the other. The major difference between the contacts of the two complexes is at Gly54I, the P_6 residue, in its interactions with Tyr104, Ala129, and Ser130.

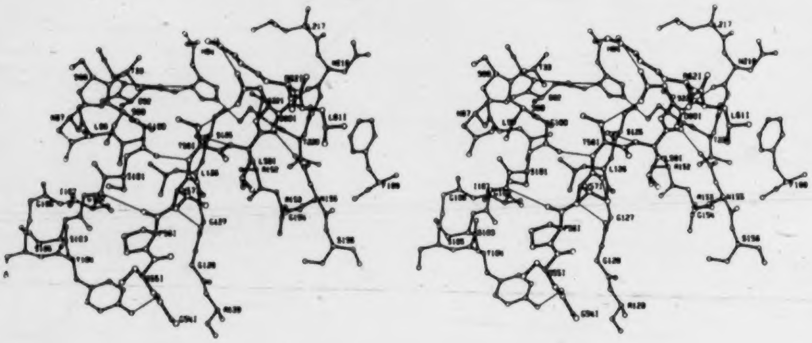


Figure 3.12. Interactions Between Subtilisin Carlsberg and Eglin-c. An ORTEP drawing (Johnson, 1965) of the active site region of subtilisin Carlsberg (open bonds) and the reactive site loop of eglin-c (filled bonds). Hydrogen bonds are indicated with thin lines. The hydrogen bond shown from the NH of Leu59I (P_1) to the carbonyl-oxygen of Ser125 is long (3.64Å) and probably weak. This bond is characteristic of an incipient tetrahedral intermediate (Robertus *et al.*, 1972). Part of the stabilization energy for the formation of a tetrahedral intermediate would be obtained by shortening, and thus strengthening, this hydrogen bond.

The S_4 specificity pocket of subtilisin Carlsberg is composed of the same residues as in subtilisin Novo, but with different orientations of some of their side-chains and different contacts to the P_4 residue, Pro56I. Changes in this region in the two complexes are discussed more fully in chapter 5. The S_1 specificity pocket changes little between the complexes, although the P_1 residues are different in the two inhibitors (Met59I in CI-2 and Leu59I in eglin-c). The remainder of the binding interactions of the residues in the reactive site loop of eglin-c are much like those in CI-2; the residues are conserved between the inhibitors (Thr58I, Arg62I) or interact primarily with other inhibitor residues (Ser55I, Asp60I).

The hydrogen-bonds between subtilisin Carlsberg and eglin-c are given in Table 3.11. The short 3-stranded β-sheet composed of residues Gly100 to Gly102, Ser125 to Gly127, and Pro56I to Leu59I is present, although Pro56I cannot form a hydrogen bond from its imide nitrogen to the carbonyl oxygen of Gly102. An additional hydrogen bond is formed between the OH of Tyr104 and the carbonyl oxygen of Gly54I, reflecting the increased proximity of these two residues in this complex. Asn155 $N^{\delta 2}$ and Ser221 NH form the hydrogen bonds in the oxyanion hole with the carbonyl oxygen of Leu59I. Two extra hydrogen bonds are made by the side-chain of the P_3' residue, Arg62I, to Asn62, that are not

Table 3.11

Hydrogen Bonds Between Subtilisin Carlsberg and Eglin-c

Site	Residues	N···O(Å)	H···O(Å)	N-H··O(°)
P_6	Tyr104 OH→Gly54I CO	2.48	1.53	158
P_4	Gly102 NH→Pro56I CO	3.40	2.54	153
P_3	Val57I NH→Gly127 CO	2.94	1.98	160
	Gly127 NH→Val57I CO	2.94	2.01	154
P_2	Thr58I NH→Gly100 CO	2.90	2.01	147
P_1	Leu59I NH→Ser125 CO	3.64	2.76	147
	Asn155 $N^{\delta 2}$→Leu59I CO	2.77	1.85	152
	Ser221 NH→Leu59I CO	3.16	2.21	158
P_2'	Leu61I NH→Asn218 CO	2.84	2.14	125
P_3'	Arg62I $N^{\eta 1}$→Asn62 CO	3.12	2.17	159
	Arg62I $N^{\eta 2}$→Asn62 CO	3.33	2.45	146

present in the subtilisin Novo:CI-2 complex although both residues are conserved.

3.4.2 The Reactive Site Bond and the Active Site

Early work on the mechanism of inhibition of α-chymotrypsin by eglin-c indicated that within 90 min. of complex formation 40% of the eglin-c had been cleaved, resulting in the release of a C-terminal fragment and the formation of a covalent bond between the enzyme and the N-terminal fragment (Seemüller *et al.*, 1980). This is akin to the inhibitory mechanism proposed for some of the plasma proteinase inhibitors (Travis & Salvesen, 1983), in which K_{hyd} is far in favor of hydrolysis. In contrast, inhibitors acting by the standard mechanism of Laskowski & Kato (1980) are hydrolyzed very slowly, and both fragments of the inhibitor are released from the enzyme. Our investigations support the view that eglin-c does act by the standard mechanism with the bacterial subtilisins, as does CI-2 (section 2.4.2).

Both eglin-c and CI-2 bind to the subtilisins in the complexes studied here in the manner of a good substrate and in a manner similar to the binding of protein inhibitors that are known to act by the standard mechanism (Fujinaga *et al.*, 1982; Marquart *et al.*, 1983). The rate constants for association of eglin-c with elastase and cathepsin G are about 10^7 M^{-1} sec^{-1} and the inhibition constants for eglin-c with the same enzymes are 10^{-11} M (Schnebli *et al.*, 1985).

At the completion of the refinement of the subtilisin Carlsberg:eglin-c complex, structure factors were calculated for the complex omitting the contribution of all atoms in the P_1 and P_1' residues. A difference electron density map with coefficients $m|F_o|-|F_c|, \alpha_c$ was calculated for the region around the active site and examined for large features. As in the subtilisin Novo:CI-2 complex, the only significant feature in this map was strong continuous electron density for the P_1 and P_1' residues. This region of the map is reproduced in Figure 3.13. The reactive site bond of eglin-c in complex with subtilisin Carlsberg is probably not cleaved to any significant extent.

The peptide of the reactive site bond in eglin-c does not show any distortion toward a tetrahedral intermediate.

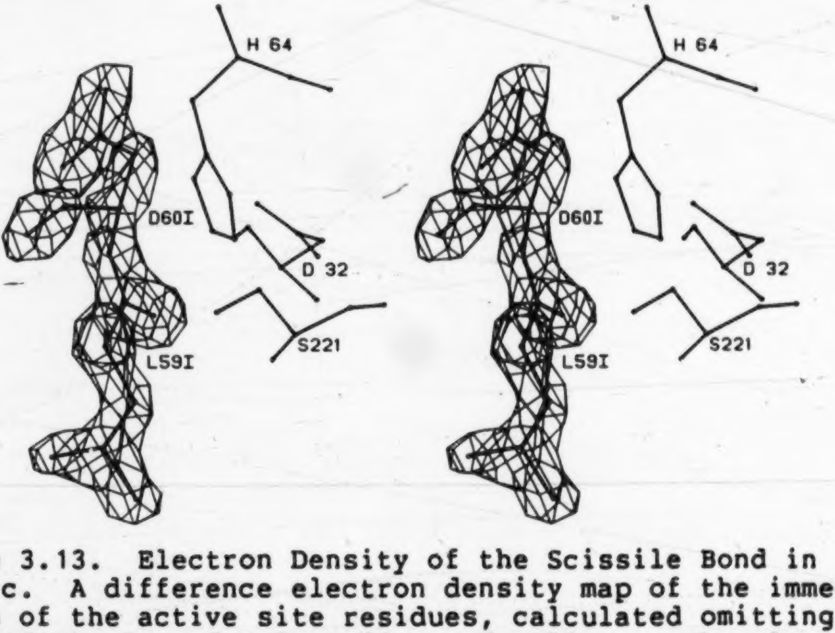


Figure 3.13. Electron Density of the Scissile Bond in Eglin-c. A difference electron density map of the immediate region of the active site residues, calculated omitting atoms of the P_1 and P_1' residues of eglin-c. The electron density of the scissile bond is strong and continuous. Contour surfaces are drawn at 0.42 e/Å³.

The θ' angle (out-of-plane bending of the carbonyl oxygen with respect to C^{α}, C, and N) of Leu59I is -4°, indicating that essentially no pyramidalization of the carbonyl carbon has taken place. The torsion angle ω (C^{α}-CO-N-C^{α}) between Leu59I and Asp60I is 170°, close to the expected value of 180° for a planar peptide bond. The mode of binding of eglin-c to subtilisin Carlsberg, the strong density of the reactive site bond, and the lack of distortion toward a tetrahedral intermediate all indicate that eglin-c acts by the standard mechanism of inhibition, not by cleavage of the reactive site bond and covalent attachment of the inhibitor N-terminus to the enzyme.

The differences in the positions of the active site residues in subtilisin Novo relative to α-chymotrypsin are also observed in subtilisin Carlsberg. The positions of the residues in the active sites of the two subtilisins are very much alike, and the interactions discussed in section 2.4.2 for the active site of subtilisin Novo are seen in subtilisin Carlsberg as well. More quantitative comparisons are made between the two active site regions of the subtilisins, as well as between the subtilisins and α-chymotrypsin, in chapter 5.

3.4.3 Solvent Structure of the Complex

The final refined model of the subtilisin Carlsberg:eglin-c complex includes 167 solvent molecules; 164 of these have been treated as water molecules and the remaining 3 have been refined as Ca^{2+} ions. The fact that both the subtilisin Novo:CI-2 and subtilisin Carlsberg:eglin-c complexes contain the same number of solvent molecules is a coincidence. Some of the better ordered and internal water molecules, and two of the Ca^{2+} ions are conserved between the complexes, but many of the water molecules are in different positions (see chapter 5). This is to be expected since most of the amino acid changes between the two enzymes

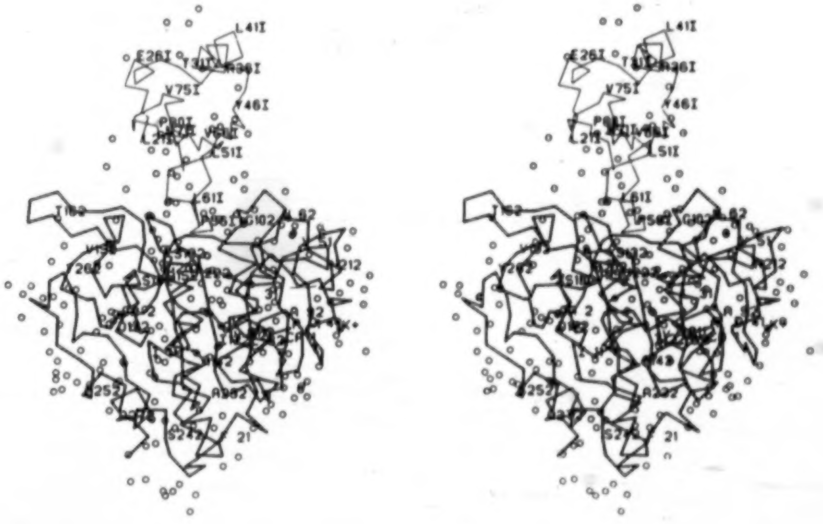


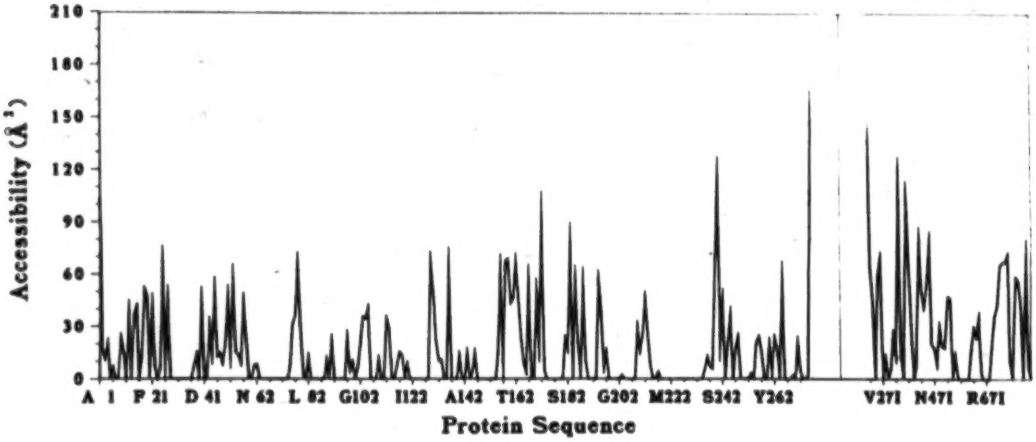
Figure 3.14. Solvent Molecules Associated With the Complex. Water molecules are indicated by open circles, Ca^{2+} and K^+ ions by closed circles, the main-chain atoms of subtilisin Carlsberg with thick lines, and the main-chain atoms of eglin-c with thin lines. All solvent molecules in the model of the crystal structure and all symmetry-related solvent molecules within 4.0Å of a protein atom have been included in this figure.

are on the molecular surface and the local properties of any particular surface region might not be strongly conserved. Different surface environments would be expected to result in different distributions of solvent molecules.

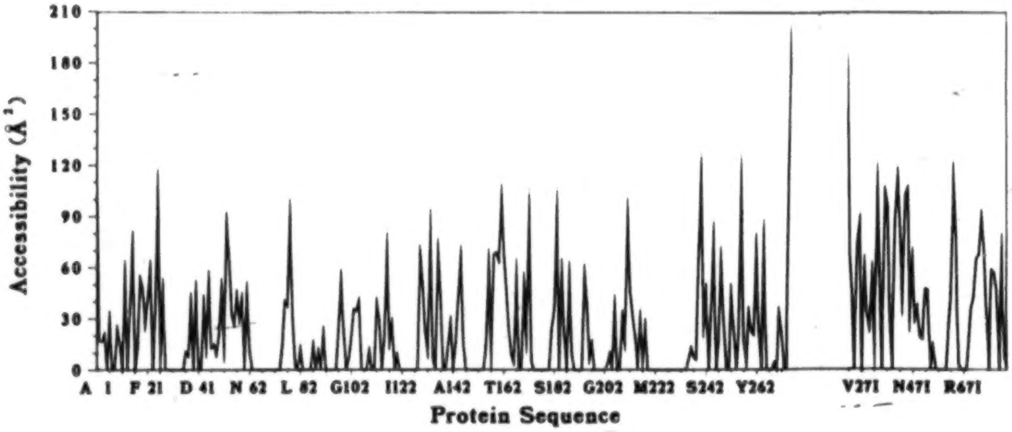
As in the subtilisin Novo:CI-2 complex, calculations of the amount of solvent expected in the unit cell indicate that only 15% of the solvent present has been modelled. All symmetry-related solvent molecules within 4.0Å of a protein atom were added in to the model for the following calculations and discussion, on the assumption that they may be considered to belong to the reference complex or be shared between neighboring molecules. There are 44 water molecules in this situation, bringing the total included in the model to 211. The subtilisin Carlsberg and eglin-c molecules are shown in Figure 3.14 with all associated solvent molecules.

The distribution of solvent molecules on the surface of the subtilisin Carlsberg:eglin-c complex and the presence of 'bare patches' were examined by the same procedure as for the subtilisin Novo:CI-2 complex (section 2.4.3). By the criterion that an atom be considered accessible if more than 10% of its total surface area is accessible in the absence of solvent or symmetry-related molecules, 15% (370) of the protein atoms in the subtilisin Carlsberg:eglin-c complex are accessible. The accessibilities of the individual amino acid residues in the complex are plotted as a function of position along the polypeptide chain in Figure 3.15.

a


210
180
150
120
90
60
30
0
Accessibility (Å²)
A 1 F 21 D 41 N 62 L 82 G102 I122 A142 T162 S182 G202 M222 S242 Y262 V271 N471 R671
Protein Sequence

b


210
180
150
120
90
60
30
0
Accessibility (Å²)
A 1 F 21 D 41 N 62 L 82 G102 I122 A142 T162 S182 G202 M222 S242 Y262 V271 N471 R671
Protein Sequence

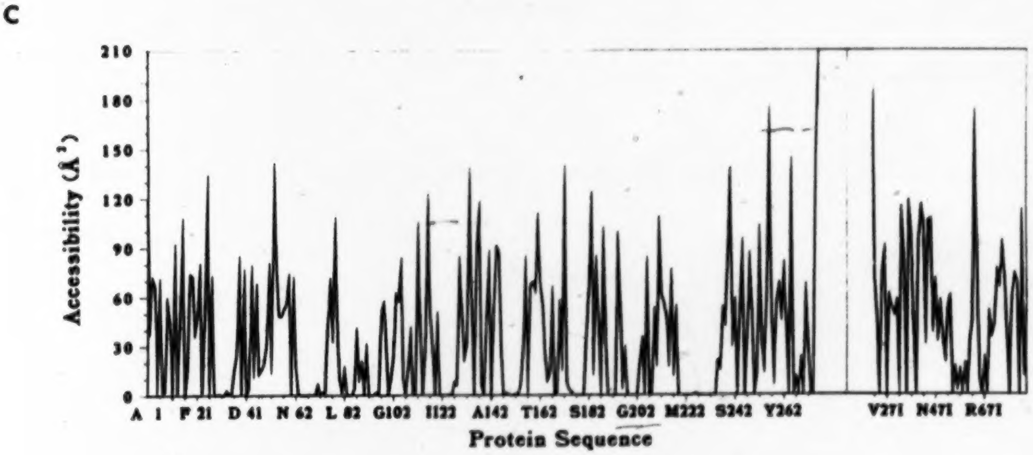


Figure 3.15. Residue Accessibilities in the Complex. Accessibilities are calculated for each atom in the complex by the method of Lee & Richards (1971). Residue accessibilities are the sum of atomic acessibilities, in Å². a) Accessibility in the presence of all solvent and symmetry-related molecules. b) Accessibility with symmetry-related protein molecules removed but all solvent within 4.0Å of the protein surface retained, including symmetry-related solvent. c) Accessibility in the absence of all solvent and symmetry-related molecules.

Sixty percent of the atoms defined as accessible do not change in accessibility if all solvent atoms are removed; these atoms constitute the 'bare patches' on the surface on the complex. Some of the bare patches are regions of contact with symmetry-related molecules. Of the atoms comprising the bare patches, 35% are in regions involved in symmetry contacts. The remaining 65% show less than 30% change in accessibility whether or not solvent and/or symmetry-related molecules are present. The subtilisin Carlsberg:eglin-c complex thus has almost twice as many atoms exposed on its surface in the environment of its crystals than the subtilisin Novo:CI-2 complex, although the two complexes have almost the same amount of ordered solvent associated with them. About 40% of the atoms in bare patches are from eglin-c. Many of the bare patches are associated with residues in turns or irregular surface structure; a few are on the hydrophilic faces of α-helices. The majority are regions of above-average temperature factor, and the solvent associated with them may also be more disordered.

Water molecules in the subtilisin Carlsberg:eglin-c complex make a total of 311 hydrogen bonds; 255 of these are to polar protein atoms and 56 to other water molecules. The hydrogen bonds between water and protein atoms can be divided into 4 groups; 18% are to main-chain nitrogen atoms, 44% to main-chain oxygen atoms, 11% to side-chain nitrogen atoms, and 27% to side-chain oxygen atoms. These percentages are almost identical to those observed for the subtilisin Novo:CI-2 complex. Higher percentages of bonds between water and oxygen atoms are seen than between water and nitrogen atoms because the vast majority of main-chain nitrogen atoms have only one ligand, whereas oxygen atoms commonly have two (Baker & Hubbard, 1984).

Of the 211 solvent molecules within 4.0Å of the surface of the subtilisin Carlsberg:eglin-c complex, 116 can be considered to be buried in the crystal environment i.e. they have accessibilities of less than 10% of their total surface area in the presence of other solvent or symmetry-related molecules. This is 55% of all the water molecules associated with the complex, close to the percentage calculated for the subtilisin Novo:CI-2 complex. All 3 of the ions bound to subtilisin Carlsberg are buried in the crystal environment, although ions 2 and 3 become marginally accessible when other solvent molecules are removed from the calculation. If all other solvent and symmetry-related solvent molecules are omitted, 49 of the buried water molecules have accessibilities of less than 10% and can be considered buried in the interior of the complex.

Several of the water molecules buried close to the active site of subtilisin Novo are conserved in subtilisin Carlsberg (section 2.4.3, Fig. 2.23). These include the channel of well-ordered water molecules between His67, Thr71, His226, and the surface of the enzyme, the water molecule bridging the center of the 5-membered turn containing Asp32, and the water molecule with hydrogen bonds to the side-chains of Asp32 and Asn123. The water molecule trapped at the bottom of the S_1 specificity pocket is also conserved.

Six water molecules associated with the inhibitor are buried in this complex, in contrast to only one in the subtilisin Novo:CI-2 complex. Three of them provide hydrogen-bonding bridges between the C-terminal strand and strand 2 of the parallel portion of the inhibitor β-sheet; one of these is conserved in the subtilisin Novo:CI-2 complex. The remaining 3 are bound to the side-chains of the two arginine residues supporting the reactive site loop; one of these is at the center of the hydrogen-bonding network stabilizing the reactive site loop.

As in the subtilisin Novo:CI-2 complex, some of the buried water molecules lie close to the surface of the complex, often covered by only a single flexible side-chain. The removal of other solvent molecules leaves 26 water molecules exposed that would be buried otherwise. There are an additional 40 water molecules that are buried in the presence or absence of other solvent, but that become exposed when symmetry-related protein molecules are removed.

3.4.4 Accessibility of Protein Residues

The variation in accessibility of individual protein residues along the polypeptide chains of subtilisin Carlsberg and eglin-c is shown in Figure 3.15. Table 3.12 gives a summary of the accessibilities for each residue type in the complex, both in the crystal environment with solvent and symmetry-related molecules present, and in their absence. The percentage of each residue type buried in the complex is also compared to the percentages found by Chothia (1975) in his survey of 9 globular proteins.

Several differences are seen in the percentages of residue types buried in the subtilisin Carlsberg:eglin-c complex compared to the figures from the survey of Chothia. Larger percentages of isoleucine, alanine, proline, histidine, and aspartic acid residues are buried in this complex; smaller percentages of phenylalanine and glutamic

Table 3.12

Accessibilities by Residue Type

Residue	No.	% Buried[1]		% Partially buried[2]		% Exposed		% Buried (Chothia)[3]
		+[4]	-	+	-	+	-	
Ile	10	100	100	0	0	0	0	65
Val	42	79	67	14	14	7	19	56
Met	5	100	100	0	0	0	0	50
Phe	8	50	25	25	0	25	75	48
Cys	0	-	-	-	-	-	-	47
Leu	20	65	40	30	20	5	40	41
Ala	41	83	63	7	2	10	35	38
Gly	38	71	45	5	11	24	44	38
Thr	23	48	26	13	0	39	74	25
Ser	34	41	18	12	9	47	73	24
Trp	1	0	0	0	0	100	100	23
Pro	15	47	40	0	7	53	53	21
Glu	6	0	0	33	0	67	100	20
His	8	62	50	0	0	38	50	19
Asp	11	36	36	18	0	46	64	15
Tyr	19	16	0	26	5	58	95	13
Asn	24	17	4	12	0	71	96	10
Gln	10	30	0	20	20	50	80	6
Lys	11	36	9	0	0	64	91	4
Arg	8	50	12	25	13	25	75	0

[1] A residue is considered buried if the sum of its atomic accessibilities is less than 5% of its potential accessible surface area in a fully extended conformation as the central residue of a Gly-X-Gly peptide (Shrake & Rupley, 1973). N- and C-terminal residues were excluded from these calculations.

[2] A residue is considered partially buried if the sum of its atomic accessibilities is between 5 and 10% of its potential accessible surface area.

[3] From a survey of 9 globular proteins (Chothia, 1975).

[4] 'Plus' is accessibility calculated with all solvent and symmetry-related molecules present, 'minus' is with no solvent or symmetry-related molecules.

acid residues are buried. Some of the differences are probably due to the small size of the sample for some residue types in the complex.

The same 4 histidine side-chains are buried in the subtilisin Carlsberg:eglin-c complex as in the subtilisin Novo:CI-2 complex: His39, His64, His67, and His226. The environment of these buried histidines is similar in the two complexes and their side-chains are fully hydrogen bonded and indirectly accessible to bulk solvent in both complexes (section 2.4.4). The nature of the hydrogen bonds of the histidine side-chains indicates that all of them but His64 are probably neutral in their protein environments. Other residues with buried and probably charged side-chains that are conserved between the two complexes, and with conserved environments, are Asp32, Asp41, Asp60, Lys94, Arg65I, and Gly83I. In the subtilisin Carlsberg:eglin-c complex, Asp60I is buried at the enzyme:inhibitor interface and involved in the hydrogen-bonding network of the reactive site loop, analogously to Glu60I in CI-2. Arg67I in eglin-c is partially buried, and involved in the same network.

The total accessible surface area of the subtilisin Carlsberg:eglin-c complex from the formulae of Chothia (1975) (section 2.4.4), based on a complex molecular weight of 35384, is 10900Å². The same area calculated by the program of J. Moult (section 2.2.4) is 11900Å² for the complex in the absence of solvent and symmetry-related molecules. If the accessible surface areas are calculated for the enzyme and the inhibitor alone, using their coordinates from the complex, the values are 9200Å² for subtilisin Carlsberg alone and 4100Å² for eglin-c alone. The difference in accessible surface area for the complex and for the sum of its two components is thus 1400Å². This value is in the range obtained for the subtilisin Novo:CI-2 complex (section 2.4.4) and for the complex of PTI with trypsin (Wodak & Janin, 1978). Assuming a free energy gain of 24 cal. for each Å² buried on formation of the complex (Chothia, 1975; Lee & Richards, 1971), the free energy gain on binding of eglin-c to subtilisin Carlsberg is 33.6 Kcal/mole.

Bibliography

Arnott, S. & Dover, S. D. (1967) *J. Mol. Biol. 30*, 209-212.

Baker, E. N. & Hubbard, R. E. (1984) *Prog. Biophys. Molec. Biol. 44*, 97-179.

Barel, A. O. & Glazer, A. N. (1968) *J. Biol. Chem. 243*, 1344-1348.

Barry, C. D., Molnar, C. E., & Rosenberger, F. U. (1976) *Technical Memo No. 229*, Computer Systems Lab, Washington University, St. Louis, MO.

Brown, I. D. & Shannon, R. D. (1973) *Acta Cryst. A29*, 266-282.

Chothia, C. (1975) *Nature 254*, 304-308.

Crawford, J. L., Lipscomb, W. N., & Schellman, C. G. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 538-542.

Creighton, T. E. (1983) *Proteins: Structures and Molecular Principles*, New York: W. H. Freeman & Co.

Crowther, A. R. (1973) in *The Molecular Replacement Method*, International Science Review *13*, ed. M. G. Rossmann, New York: Gordon & Breach, 173-178.

Cruickshank, D. W. J. (1949) *Acta Cryst. 2*, 65-82.

Cruickshank, D. W. J. (1954) *Acta Cryst. 7*, 519.

Cruickshank, D. W. J. (1967) in *International Tables for X-ray Crystallography* Vol. 2, eds. J. S. Kasper & K. Lonsdale, Birmingham: Kynoch Press, 318-340.

Davies, D. R. & Segal, D. M. (1971) in *Meth. Enzymol.* Vol. XXII, ed. W. B. Jacoby, New York: Academic Press, 266-269.

Einspahr, H. & Bugg, C. E. (1981) *Acta Cryst. B37*, 1044-1052.

Einspahr, H. & Bugg, C. E. (1984) in *Metal Ions in Biological Systems* Vol. 17, ed H. Sigel, New York: Marcel Dekker, 51-97.

Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M. Jr., & James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868-4872.

Grütter, M. G., Märki, W., & Walliser, H.-P. (1985) *J. Biol.*

Chem. 260, 11436-11437.

Hendrickson, W. A. & Konnert, J. H. (1980) in *Biomolecular Structure, Function, Conformation and Evolution* Vol. I, ed. R. Srinivasan, Oxford: Pergamon Press, 43-57.

Jacobs, M., Eliasson, M., Uhlén, M., & Flock, J.-I. (1985) *Nucleic Acids Research 13*, 8913-8926.

James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol. 163*, 299-361.

Johnson, C. K. (1965) ORTEP, Report ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, TN.

Kabsch, W. & Sander, C. (1983) *Biopolymers 22*, 2577-2637.

Keay, L. & Moser, P. W. (1969) *Biochem. Biophys. Res. Commun. 34*, 600-604.

Laskowski, M. Jr. & Kato, I. (1980) *Ann. Rev. Biochem. 49*, 593-626.

Lee, B. & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379-400.

Linderstrøm-Lang, K. & Ottesen, M. (1947) *Nature 159*, 807-808.

Luzzati, V. (1952) *Acta Cryst. 5*, 802-810.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Cryst. B39*, 480-490.

Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491-497.

McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242-7246.

McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, New York: John Wiley & Sons, 96-97.

Melville, J. C. & Ryan, C. A. (1972) *J. Biol. Chem. 247*, 3445=3453.

North, A. C. T., Phillips, D. C., & Mathews, F. S. (1968) *Acta Cryst. A24*, 878-884.

Ramachandran, G. N. & Mitra, A. K. (1976) *J. Mol. Biol. 107*, 85-92.

Ramakrishnan, C. & Ramachandran, G. N. (1965) *Biophys. J. 5*, 909-933.

Rao, S. N., Jih, J.-H., & Hartsuck, J. A. (1980) *Acta Cryst. A36*, 878-884.

Read, R. J. (1986) *Acta Cryst.*, in press.

Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420-4433.

Rink, H., Liersch, M., Sieber, P., & Meyer, F. (1984) *Nucleic Acids Research 12*, 6369-6387.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., & Wilcox, P. E. (1972) *Biochemistry 11*, 2439-2449.

Rossmann, M. G. (1973) *The Molecular Replacement Method*, International Science Review 13, New York: Gordon & Breach.

Schnebli, H. P., Seemüller, U., Fritz, H., Maschler, R., Liersch, M., Bodmer, J. L., Virca, G. D., Lucey, E. C., Stone, P. G., & Snider, G. L. (1985) in *Intracellular Protein Catabolism*, eds. E. A. Khairallah, J. S. Bond, & J. W. C. Bird, New York: Alan Liss, Inc., 287-290.

Seemüller, U., Meier, M., Ohlsson, K., Müller, H.-P., & Fritz, H. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1105-1117.

Seemüller, U., Eulitz, M., Fritz, H., & Strobl, A. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1841-1846.

Shrake, A. & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351-372.

Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1979) *J. Mol. Biol. 134*, 781-804.

Sielecki, A. R., James, M. N. G., & Broughton, C. G. (1982) in *Computational Crystallography*, ed. D. Sayre, Oxford: Oxford University Press, 409-419.

Smith, E. L., DeLange, R. J., Evans, H., Landon, M., & Markland, F. S. (1968) *J. Biol. Chem. 243*, 2184-2191.

Svendsen, I., Boisen, S. & Hejgaard, J. (1982) *Carlsberg Res. Commun. 47*, 45-53.

Thiessen, W. E. & Levy, H. A. (1973) *J. Appl. Cryst. 6*, 309.

Travis, J. & Salvesen, G. S. (1983) *Ann. Rev. Biochem. 52*, 655-709.

Wodak, S. J. & Janin, J. (1978) *J. Mol. Biol. 124*, 323-342.

Wright, C. S., Alden, R. A., & Kraut, J. (1969) *Nature (London) 221*, 235-242.

4. Native CI-2[1]

CI-2 is a member of the potato inhibitor 1 (PI-1) family of serine proteinase inhibitors in the classification of Laskowski & Kato (1980). It was isolated partly on the basis of its activity as an inhibitor of α-chymotrypsin and microbial serine proteinases, and assigned to the PI-1 family by sequence comparisons (Svendsen *et al.*, 1980b). CI-2 is the only member of this family for which the structure of the native inhibitor, crystallized in the absence of an enzyme, is known. Inhibitors from other families that have had their X-ray structures solved in the native state are pancreatic trypsin inhibitor (PTI, reviewed in Marquart *et al.*, 1983) from the Kunitz family, the Japanese quail (Papamakos *et al.*, 1982) and silver pheasant (Bode *et al.*, 1985) third domains of ovomucoid inhibitors from the Kazal family, *Streptomyces* subtilisin inhibitor (SSI, Mitsui *et al.*, 1979) and plasminostreptin (PS, Kamiya *et al.*, 1984) from the SSI family, and α_1-proteinase inhibitor (Loebermann *et al.*, 1984) from the α_1-proteinase inhibitor family.

Crystals of CI-2 in the native state have been grown. Initial attempts were made to solve the structure by the multiple isomorphous replacement method (Bragg & Perutz, 1954; Blow & Crick, 1959; Matthews, 1974), but were unsuccessful. The structure was subsequently solved by the molecular replacement method (Rossmann, 1973). The complete polypeptide backbone of the inhibitor could be traced in the initial electron density map, beginning at Lys21I. The least-squares refinement of CI-2 was completed after 133 cycles, at an *R*-factor of 0.198 for data in the resolution range 8.0 to 2.0Å. The secondary and tertiary structures of the inhibitor have been examined. The structure of the native CI-2 is compared to that of CI-2 in complex with subtilisin Novo in chapter 5.

---

[1] A version of portions of this chapter has been published [McPhalen, C. A., Evans, C., Hayakawa, K., Jonassen, I., Svendsen, I., & James, M. N. G. (1983) *J. Mol. Biol.* *168*, 445-447].

4.1 Structure Solution and Refinement

4.1.1 Crystallization

Purified lyophilized CI-2 was prepared at the Carlsberg Research Center, Copenhagen, Denmark (Jonassen, 1980; Svendsen *et al.*, 1980a). During purification procedures, the inhibitor is subject to hydrolysis at 3 bonds near its amino-terminus (following Asn11I, Gly15I, and Arg17I, Fig. 2.2), and the product with the 'ragged' amino-terminus was used in crystallization trials. The intact inhibitor contains 83 amino acids, M_r = 9250 (Svendsen *et al.*, 1980a) (Fig. 2.2). Small crystals of native CI-2 were grown by the hanging-drop vapor-diffusion method (Davies & Segal, 1971; McPherson, 1982). CI-2 (2.0 mg/mL) was dissolved in distilled water and a hanging drop was prepared consisting of equal amounts of the protein solution and a buffer of 40% (w/v) $(NH_4)_2SO_4$ and 50 mM Tris·HCl at pH 8.0. The reservoir of the crystallization tray was filled with 1.0 mL of buffer solution. Small hexagonal prisms grew after one to two months, but these crystals were plate-like and less than 0.1 mm thick in one dimension. Subsequent seeding produced crystals of sufficient size for high resolution X-ray analysis, 3 to 4 weeks after placing small crystals into freshly prepared hanging drops of protein and buffer solutions.

Precession photographs show that the crystals diffract to a minimum *d* spacing of 2.0Å. There are no systematic absences. The lattice symmetry is consistent with space group P622. The crystals have unit cell dimensions of $a = b = 69.1(1)$Å, $c = 52.9(1)$Å, and $\gamma = 120°$. The volume per unit molecular weight, V_m, is 2.0Å^3/dalton, consistent with one protein molecule per asymmetric unit and a solvent content of 38% (Matthews, 1968).

4.1.2 Data Collection and Processing

Intensity data sets for the native inhibitor were collected from 3 crystals, described in Table 4.1. A number of data sets were also collected for a variety of potential heavy atom derivatives, discussed in section 4.1.3. All data sets were collected on a Nonius CAD4 diffractometer, with incident radiation Ni-filtered CuK$_\alpha$ at 40 kV and 26 mA, and a 1.3 mm diameter incident beam collimator. The crystal-to-counter distance was 60 cm, and the diffracted beam passed through a He-filled tunnel. The size of the peak scan varied among the data sets from 0.5° to 0.66° in

Table 4.1

Crystal Data for Native CI-2

Crystal	1	2	3
Resolution range (Å)	39.7-2.8	21.0-2.0	59.5-2.0
Total reflections measured	4283	11362	12197
Total unique reflections	2075	5597	5903
Total reflections $I \geq \sigma(I)$[1]	1625	3783	4471
R_{merge}[2]	0.062	0.043	0.046
Maximum absorption correction	1.36	1.33	1.86
Maximum decay correction	4 %	9 %	9 %
Scale	30.49	19.72	13.93
Overall B (Å^2)	23	21	21

[1] $\sigma(I) = (I + c^2 I^2 + (t_I/t_{Bk})^2 (\Sigma Bk + c^2 \Sigma Bk^2))^{1/2}$, where I = total intensity; Bk = background counts; t_I = time for intensity measurement; t_{Bk} = time for background measurement; c = instrument instability constant = 0.01.

[2] $R_{\mathrm{merge}} = \Sigma_{hkl}(\Sigma_i |I_i - \langle I \rangle| / \Sigma_i I_i)$ for reflections measured more than once in a data set.

ω; the scan rate was constant at 0.67°/min. Backgrounds were scanned for 1/4 of the width of the peak scan in ω on either side of the peak scan. The data for both 2.0Å data sets were collected in shells to minimize the loss of weak reflections at higher resolution. Two shells were collected; first from 2.0 to 2.5Å, and then from 2.5Å to the lower resolution limit of each data set (Table 4.1).

During the processing of each data set, the measured backgrounds were corrected for intensity dependence and an averaging function of 2θ and ϕ was applied. The absorption correction applied was an empirical function of ϕ (North *et al.*, 1968). Lorentz and polarization corrections were also applied. The absolute scale and an overall temperature factor (B, Å^2) were calculated with the program ORESTES (Thiessen & Levy, 1973). Decay due to radiation damage was monitored by the periodic collection of 5 standard reflections; the decay correction applied to the data was an empirical function of the decay of these standards with time.

4.1.3 Structure Determination

The initial approach to solving the structure of CI-2 was to use the isomorphous replacement method (Bragg & Perutz, 1954; Blow & Crick, 1959; Matthews, 1974). The general procedure followed was to soak medium-sized crystals of CI-2 in solutions of heavy metal salts or organometallic compounds at concentrations of 1 to 10 mM, for periods ranging from 24 hours to 2 weeks. The heavy metal compounds tested were chosen on the basis of their ability to scatter X-rays strongly. Precession photographs were taken of the crystals after soaking and were compared by eye to precession photographs of CI-2 crystals that had not been soaked. If the photographs of the soaked crystals showed noticeable changes in the intensities of the diffraction pattern, a full set of diffraction data was collected on the diffractometer for that heavy atom derivative.

Ideally, the heavy atoms should be incorporated at a few specific sites in the asymmetric unit of the crystal and should contribute strongly to the diffraction pattern, but should not change the structure of the protein significantly. If these conditions are fulfilled, a Patterson map can be calculated with coefficients that are equal to the difference in the intensities of the heavy atom derivative and native protein diffraction patterns ($|F_{PH}| - |F_P|$). Such a map will show the interatomic vectors between all the heavy atoms in the unit cell, and the positions of the heavy atoms may be determined from it. If the positions of the heavy atoms are known, their contribution to the structure factor phases and intensities of the diffraction pattern from the derivative crystals may be calculated. With this knowledge, the phases of the protein contribution to the diffraction pattern may be determined and combined with the intensities from the diffraction pattern of the native protein to calculate an electron density map of the native protein.

The heavy atom compounds tested that gave noticeable changes from the diffraction pattern of the native protein crystals were $Sm_2(SO_4)_3 \cdot 8H_2O$, mersalate sodium mersalyl (an organo-mercury compound), K_2PtCl_4, $UO_2(Ac)_2$, K_2HgI_4, and $AuCl_4$. Peaks were present in all the Patterson maps, but the protein phases determined from each of the derivatives were not consistent, indicating that the true positions of the heavy atoms had not been found in the Patterson maps.

The structure of CI-2 was subsequently solved by the molecular replacement method (see below) and refined at 2.0Å resolution. Difference electron density maps were then calculated from the intensities of the native and heavy atom derivative structure factors, and the phases from the refined native structure ($(|F_{PH}| - |F_P|, \alpha_P)$). In one case, the K_2PtCl_4 derivative, the peaks on the difference map showed little correspondence to those from the Patterson map. One cell dimension of these crystals changed by 1.5% upon soaking; this change may have resulted in sufficient non-isomorphism of native and derivative crystals to invalidate the use of the isomorphous replacement method. For all the other derivatives, the major heavy atom sites seen in the difference maps were very close to special positions on the symmetry axes of the crystal cell. The side-chain of the single methionine residue in the inhibitor, Met59I, lies close to a two-fold axis; most of the heavy atom derivatives were expected to bind to this residue, and it appears as though they actually did to some extent. Unfortunately, the computer programs used to find the phases of the native protein from the heavy atom positions seem to have been unable to handle positions so close to a symmetry axis.

Once the structure of CI-2 in complex with subtilisin Novo was known, the method of molecular replacement was used to solve the phase problem for the native CI-2 (Rossmann, 1973). The search model was the crystal structure of CI-2 from the complex with subtilisin Novo, partially refined to an *R*-factor of 0.193 at 2.1Å resolution. Data from crystal 2 (Table 4.1) were used for the molecular replacement calculations. The procedures and choice of parameters in the molecular replacement strategy used here are similar to those discussed in section 2.1.3.

The rotation search was performed with the fast rotation function of Crowther (1973) (program of E. Dodson); the search maximized the agreement between the normalized structure factors ($|\mathbf{E}|$'s) of the model and the native CI-2 data. The $|\mathbf{E}_c|$'s were calculated for the CI-2 model placed in an orthogonal unit cell of symmetry P1, with each edge of the cell measuring 42Å. The $|\mathbf{E}_o|$'s were calculated for the unknown structure, from the structure factors derived from data processing, with the program ORESTES (Thiessen & Levy, 1973). The Patterson maps calculated from these two sets of $|\mathbf{E}|$'s were compared between 4.0 and 18.0Å from the origin. The resolution range of reflections included in the search was 10.0 and 3.0Å. The 2292 $|\mathbf{E}_c|$'s with magnitudes greater than 1.00 were used in the search: 1301 $|\mathbf{E}_o|$'s with magnitudes greater than 0.10 were included.

The initial coarse rotation function search was performed on an asymmetric unit of rotation function space (Rao *et al.*, 1980), on a grid of 5° in each of the Eulerian angles α, β, and γ (Crowther, 1973). The maximum peak in the rotation function map was 5.6σ above the mean and 2.3σ above the next highest peak. This rotation function solution was not as dramatically correct as those seen for the subtilisin Novo:CI-2 and subtilisin Carlsberg:eglin-c complexes, but the size of the peak and its height above other peaks were sufficient for it to be unambiguous. A second rotation search was performed with 1° intervals in β and 5° intervals in α and γ about the maximum of the coarse search. The fine search gave a maximum peak of 4.8σ above the mean of the second map. The final values of the orientation parameters from the rotation search were interpolated from the fine search map to 0.2° in β and 1° in α and γ.

The translation function search area for the native CI-2 molecule, in space group P622, is $0 \leq x \leq 1$, $0 \leq y \leq 1$, and $0 \leq z \leq 1/2$. Data between 4.0 and 5.0Å were used for the translation function search, performed with the local program BRUTE, written by M. Fujinaga. The grid spacing for the initial coarse search was set at 1.1Å. The search gave a peak of 4.8σ above the mean of the map and 1.1σ above the next highest peak. A second search was performed about this peak at grid intervals of 0.2Å, extending ±1.0Å along each axis. The second search also varied the rotational parameters at intervals of 3°, extending ±3° about the angular values from the rotation search. To refine the complete molecular replacement solution further, two fine searches were performed at grid intervals of 0.1Å in the translation component and 1° in the rotation component. The second fine search was offset from the first by 0.5° in α to define the location of the peak maximum more closely. The fine searches gave a single peak of 5.9σ above the mean, a significant improvement from the initial coarse search. The improvement was due mainly to better definition of the rotational parameters; the value of α for the final solution was 3° smaller than that found by the rotation function search.

The *R*-factor for the transformed coordinates of the model of CI-2 from the complex with subtilisin Novo was 0.50 for all data to 2.0Å resolution. This model was used in the calculation of structure factor amplitudes and phases for the computation of the initial electron density map. As for the structure solutions of the two complexes discussed previously, all electron density maps were calculated with the coefficients of Read (1986), designed to suppress model bias resulting from phasing by partial structures with errors (section 2.1.3).

4.1.4 Structure Fitting and Refinement

Throughout the fitting and refinement of the native CI-2 structure, the MMS-X interactive graphics (Barry *et al*., 1976) with the macromolecular modelling system M3, developed by C. Broughton (Sielecki *et al*., 1982), was used for map interpretation and model fitting.

The initial electron density map from the molecular replacement solution was of good quality. No large conformational changes were seen in the CI-2 molecule, although the molecular replacement model required minor adjustment in

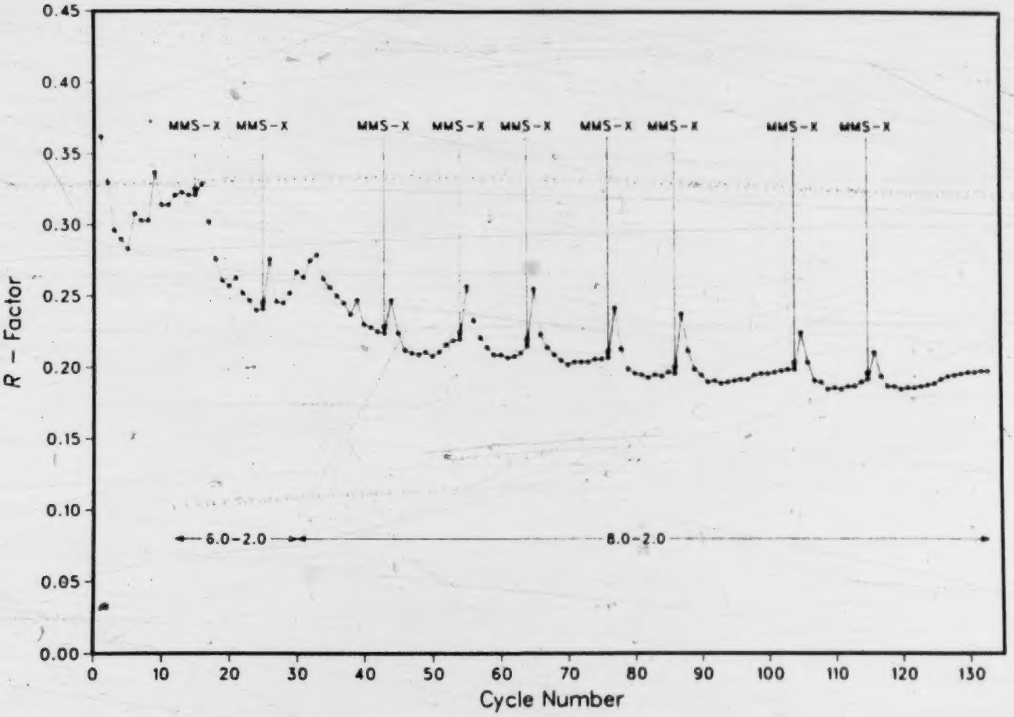


Figure 4.1. Progress of CI-2 Refinement. The *R*-factor at each cycle is plotted versus cycle number. The resolution range of the data included at each cycle is indicated. Data from 6.0 to 2.8Å were used for cycles 1 to 5, from 6.0 to 2.5Å for cycles 6 to 8, and 6.0 to 2.2Å for cycles 9 to 11. Electron density maps with coefficients $2m|F_o|-D|F_c|$, α_c were calculated and used to refit the model on the graphics system at the points marked "MMS-X".

most regions. The model of CI-2, adjusted to the initial electron density map, was used to begin least-squares refinement.

The restrained-parameter least-squares refinement program of Hendrickson & Konnert (1980), modified locally by M. Fujinaga for the FPS164 attached processor, was used for the refinement of the atomic parameters of CI-2. The parameters that were restrained and the general refinement strategy are the same as those used for the subtilisin Novo:CI-2 complex (section 2.1.4). The course of the

Table 4.2

Course of Least-squares Refinement

Cycle number	Data used(Å)	Number of reflections	R-factor	Comments
1	6.0-2.8	1610	0.362	Start refinement with model containing residues Leu20I to Gly83I. Include all reflections with $I \geq 2\sigma(I)$. One overall *B*-factor.
6	6.0-2.5	2171	0.308	Add data from 2.8 to 2.5Å.
9	6.0-2.2	2954	0.337	Add data from 2.5 to 2.2Å.
12	6.0-2.0	3506	0.321	Add data from 2.2 to 2.0Å.
15	6.0-2.0	3506	0.321	Refit inhibitor on graphics system.
16	6.0-2.0	3506	0.329	Restart refinement with individual atomic *B*-factors.
25	6.0-2.0	3506	0.241	~~Refit inhibitor~~. Calculate difference map, coefficients $m\lvert F_o\rvert - \lvert F_c\rvert$, α_c. Choose solvent positions from difference map peaks. Add Asn19I at N-terminus.
26	6.0-2.0	4177	0.276	Restart refinement with residues Asn 19I to Gly83I, 22 H_2O. Include all reflections with $I \geq \sigma(I)$.
30	8.0-2.0	4308	0.267	Add data from 8.0 to 6.0Å.
43	8.0-2.0	4308	0.224	Refit inhibitor, choose new H_2O.
44	8.0-2.0	4308	0.247	Restart refinement with 35 H_2O.
54	8.0-2.0	4308	0.220	Refit inhibitor, choose new H_2O.
55	8.0-2.0	4308	0.257	Restart refinement with 49 H_2O. Remove Thr58I and Met59I from refinement - poor density in region of reactive site.
64	8.0-2.0	4308	0.215	Refit inhibitor, choose new H_2O. Add Thr58I and Met59I.
65	8.0-2.0	4471	0.255	Restart refinement with 61 H_2O. Use data from crystal 3 (Table 4.1).
71	8.0-2.0	4471	0.204	Delete H_2O with occupancies less than 0.2. Restart refinement with 49 H_2O.

(Table 4.2 continued)

76	8.0-2.0	4471	0.207	Refit inhibitor, choose new H_2O.
77	8.0-2.0	4471	0.242	Restart refinement with 69 H_2O.
86	8.0-2.0	4471	0.196	Refit inhibitor and H_2O. Remove H_2O with weak density, choose new H_2O from difference map.
87	8.0-2.0	4471	0.238	Restart refinement with 72 H_2O.
104	8.0-2.0	4471	0.199	Refit inhibitor, remove poor H_2O and choose new H_2O from difference map.
105	8.0-2.0	4471	0.225	Restart refinement with 71 H_2O.
115	8.0-2.0	4471	0.192	Refit inhibitor. Remove H_2O with contacts of less than 2.0Å to symmetry-related H_2O.
116	8.0-2.0	4471	0.211	Restart refinement with 64 H_2O.
133	8.0-2.0	4471	0.198	End of refinement.

refinement is summarized in Table 4.2 and the changes in the *R*-factor during the refinement are shown in Figure 4.1. Data from crystal 2 were used for the first 64 cycles, and data from crystal 3 for cycles 65 to 133 (Table 4.1). Most regions of the native CI-2 molecule were similar enough to the molecular replacement model from the complex with subtilisin Novo that large manual adjustments during refitting were not necessary. Density in the region of the reactive site bond (Met59I - Glu60I) was poor throughout the refinement and the *B*-factors of these residues are high.

No density was seen for residues 1I to 18I in the electron density maps at any stage of refinement, and the tight packing of the surrounding symmetry-related protein molecules indicated that no large extra segment of CI-2 was disordered but present in the crystals. The size of the inhibitor in the crystal was compared to that of the inhibitor in the lyophilized purified powder provided by the Carlsberg Laboratories by SDS polyacrylamide gel electrophoresis. The gel is shown in Figure 4.2. The major band in each of lanes 2 and 3, running at a molecular weight of approximately 6500 dalton, is CI-2 (both CI-2 and eglin-c are known to run at anomalously low molecular weights on SDS gels (Seemüller *et al.*, 1981; Svendsen *et al.*, 1982)). Although the bands are broad, the major component of CI-2 from the crystals is clearly running at a lower molecular weight than that of the lyophilized powder, indicating that some further hydrolysis has taken place during the

1
2
3
43k
26k
18k
14k
6k
3k

Figure 4.2. SDS Gel of CI-2 from Crystals. A 15% Laemmli polyacrylamide gel containing 0.1% SDS and 6M urea (gel run by B. Lemire). Lane 1 contains molecular weight standards: ovalbumin (43000), α-chymotrypsinogen (25700), β-lactoglobulin (18400), lysozyme (14300), bovine trypsin inhibitor (6200), and insulin (A and B chains, 3000). Lane 2 contains CI-2 from crushed and redissolved crystals. Lane 3 contains CI-2 from the lyophilized purified powder used to grow crystals. The gel was stained with Coomassie Brilliant Blue.

crystallization procedure. Some minor bands can be seen in both lanes 2 and 3 at a lower molecular weight than the major bands; these are probably some of the cleavage products. The molecular weight of these products cannot be determined accurately from a single SDS gel, thus the exact point of cleavage of the protein in the major band cannot be calculated. The conclusion drawn from this experiment is that part of the N-terminus of CI-2 is not present in the crystalline protein, so it is not surprising that it cannot be seen in the electron density map. The size of the cleaved portion of the N-terminus relative to the size of the disordered portion in the crystals is not known. The faint bands in lanes 2 and 3 that run at molecular weights higher than the major form of CI-2 may be aggregation products or minor contaminants.

4.1.5 Quality of the Refined CI-2 Structure

The refinement of the native CI-2 structure was halted at cycle 133. The model had essentially stopped changing, although the R-factor was still high at this stage compared to the fully refined complex structures of chapters 2 and 3. This section describes the criteria used to judge the completeness of the refinement and the quality of the refined structure.

The final refinement parameters are given in Table 4.3. The restraints on the final structure are tight and are the same as those at the final refinement cycles of the two complexes; the indicated parameter shifts at the final cycle are small. The rms coordinate shift is 0.014Å; the maximum coordinate shifts are 0.068Å for the N^{ζ} of Lys37I in the protein and 0.102Å for solvent molecules. The final parameter shifts for this molecule are similar in size to those of the subtilisin:inhibitor complexes.

Table 4.3

Final Refinement Parameters and Results

No. of cycles	133
R-factor	0.198
Resolution range (Å)	8.0-2.0
No. of reflections ($I \geq \sigma(I)$)	4471
No. of protein atoms	521
No. of solvent atoms	64
No. of variable parameters	2405
$<\|\|\mathbf{F}_o\|-\|\mathbf{F}_c\|\|>$	40
<Coordinate shift> (Å)	0.012
<*B*-factor shift> (Å^2)	0.25
Rms deviations from ideal values[1]	
Distance restraints (Å)	
Bond distance	0.007(0.008)
Angle distance	0.029(0.016)
Planar 1-4 distance	0.021(0.016)
Plane restraint (Å)	0.014(0.012)
Chiral-center restraint (Å^3)	0.123(0.080)
Non-bonded contact restraints (Å)	
Single torsion contact	0.297(0.400)
Multiple torsion contact	0.276(0.400)
Possible hydrogen bond	0.240(0.400)
Isotropic thermal factor restraints (Å^2)	
Main-chain bond	1.956(2.000)
Main-chain angle	2.909(2.000)
Side-chain bond	3.611(3.000)
Side-chain angle	5.285(3.000)

[1] The values of σ, in parentheses, are the input estimated standard deviations that determine the relative weights of the corresponding restraints [see Hendrickson and Konnert (1980)].

The refinement of the native CI-2 structure proceeded less smoothly than those of the two complexes. If the full shifts in atomic parameters indicated at each cycle were applied to the model, the refinement tended to diverge, rather than converging. This problem was overcome by applying only half the indicated shifts to all atomic parameters, beginning at cycle 39. Due to the lower resolution and higher overall *B*-factor for the native CI-2, the electron density of the inhibitor was generally less well-defined than that of the complexes. This meant that fitting a model to the density with good stereochemistry was more difficult, and proved to be almost impossible in some regions e.g. the two N-terminal residues, Asn19I and Leu20I, and at the reactive site bond, Met59I and Glu60I. In spite of such difficulties in the manual fitting of the model on the graphics system, the good geometry of the final refined structure has not been achieved at the expense of good agreement between observed and calculated structure factors. The *R*-factor and the mean overall difference between the observed and calculated structure factors rose little as the restraints on geometry were tightened in the final refinement cycles (Fig. 4.1).

A difference electron density map was calculated after the final refinement cycle and examined on the graphics system. The highest peak on the difference map was 0.48e/Å^3, and there were 16 other peaks with heights greater than 0.40e/Å^3. Ten of the peaks were located on symmetry axes; the other 7 were too small and poorly defined to be considered as additional solvent molecules. There were 29 negative troughs in the difference map with depths greater than 0.40e/Å^3. Fourteen of the troughs are associated with segments of the model that are not centered properly in the electron density; the density does not conform to a model with good geometry. The remaining troughs are close to

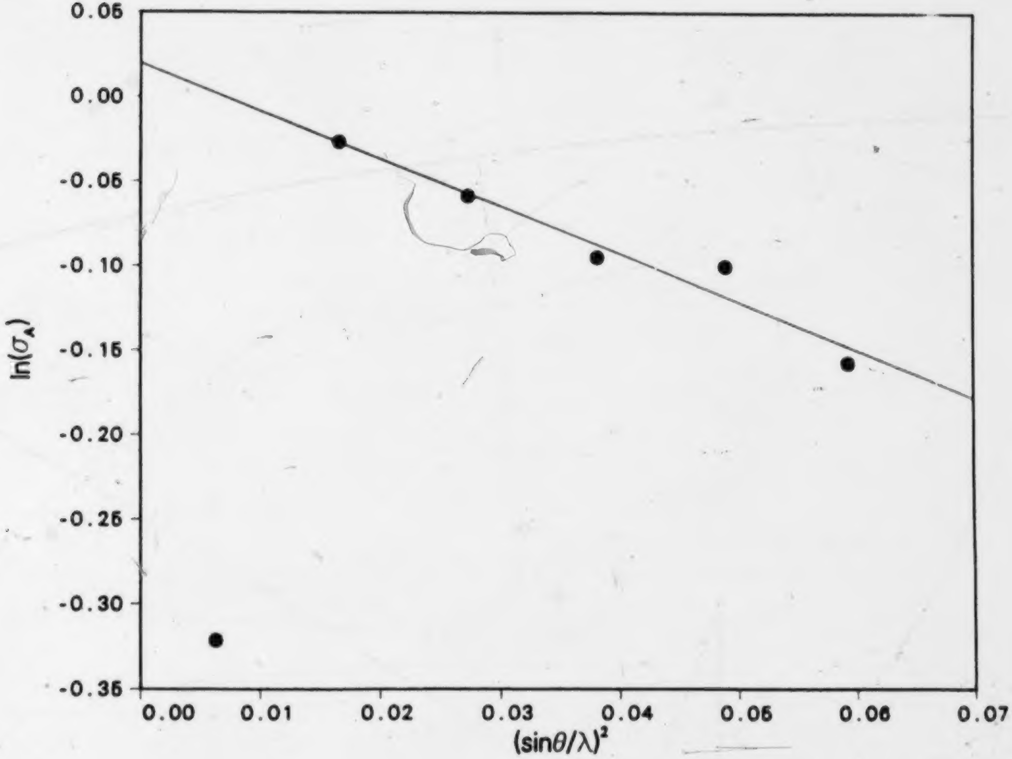


Figure 4.3. σ_A Plot to Estimate Coordinate Error. The slope of the straight-line portion of this plot is $[-\pi^2(<|\Delta r|>)^2]$ (mean coordinate error), or $[-(8\pi^2/3)<|\Delta r|^2>]$ (rms coordinate error). The line is determined by a least-squares fit to all data points but the first one. The first data point describes low resolution data that is affected by omission of bulk solvent from the protein model.

atoms with relatively high *B*-factors.

The accuracy of the atomic coordinates in the refined structure of the native CI-2 was estimated with two methods: the σ_A plot of Read (1986) and the method of Cruickshank (1949, 1954, 1967) (section 2.1.5). The σ_A plot for the native CI-2 is given in Figure 4.3. The slope of the straight-line portion of the plot is -2.374, corresponding to a mean coordinate error of 0.28Å and an rms coordinate

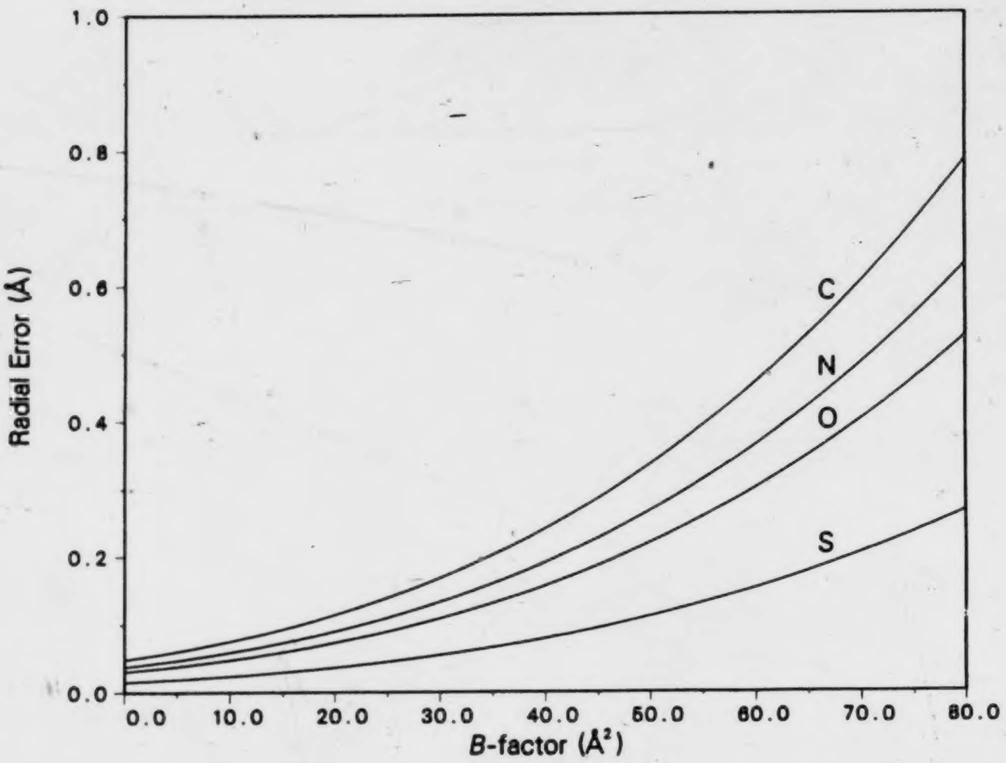


Figure 4.4. Atomic Coordinate Error by the Method of Cruickshank. The estimated radial standard deviations in atomic position are given as a function of *B*-factor. The 4 curves, from top to bottom, are for carbon, nitrogen, oxygen, and sulfur atoms in the final refined structure of CI-2.

error of 0.30Å. These errors are close to those estimated for the subtilisin Novo:CI-2 complex refined in a similar resolution range, but the complex has a much lower final *R*-factor.

The individual atomic coordinate errors may be estimated with the method of Cruickshank. The space group P622 was treated as monoclinic, with the non-orthogonal angle equal to 120°. The coordinate error along the z axis was calculated with the Cruickshank formula for a

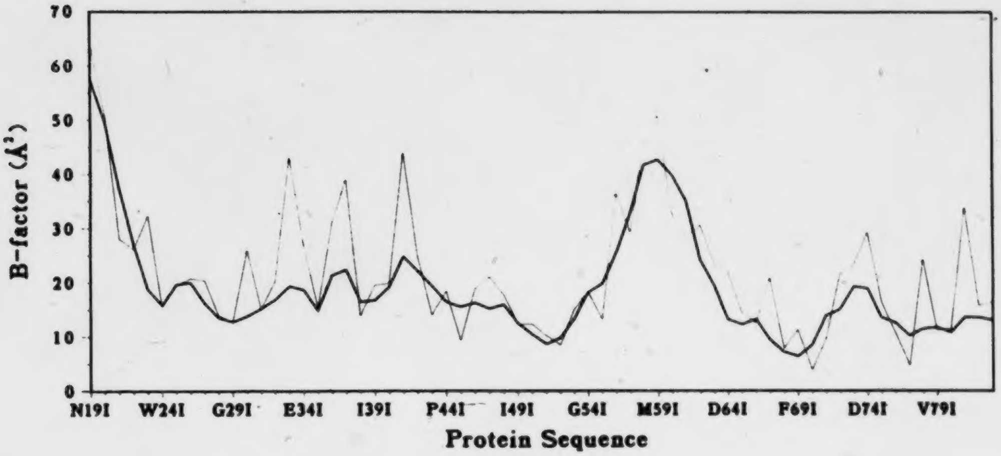


Figure 4.5. Variation in *B*-factor Along the Polypeptide Chain. The heavy line denotes the mean *B*-factor of the main-chain atoms, the light line that of the side-chain atoms. No *B*-factors are given for residues 1I-18I of the inhibitor; they are not seen in the electron density map. The mean overall *B*-factor for the atoms of the structure is 21Å².

non-centered monoclinic space group (section 2.1.5); in the monoclinic case, $\sigma(z) = \sigma(x) \simeq \sigma(y)$. Radial errors were calculated for each atom type in the structure and for a range of *B*-factors, with $\overline{\sigma}(r_i) = \sqrt{3}\sigma(z_i)$. A plot of radial error as a function of *B*-factor is given in Figure 4.4 for the carbon, nitrogen, oxygen, and sulphur atoms in CI-2. An estimate of the accuracy of a particular region of the structure can be obtained using the data of Figure 4.4 with Figure 4.5, showing *B*-factors as a function of position along the polypeptide chain. Two features of Figure 4.5 are of particular interest. The reactive site loop of the native inhibitor (residues 56 to 62) has very high *B*-factors, second only to the disordered residues at the

N-terminus. As well, a distinctive 'sawtooth' pattern can be seen for *B*-factors of residues in the α-helix (31 to 43). Residues in positions on the outside of the helix, exposed to bulk solvent, have *B*-factors corresponding to peaks in the pattern, and residues on the inside correspond to troughs. Similar patterns can be seen less clearly for the amphipathic helices in the subtilisins (Fig. 2.8, Fig. 3.4).

The rms value of the estimated radial errors for all the atoms in the model is an estimate of the overall coordinate error in the structure. This value for the native CI-2 is 0.24Å, close to the overall rms error of 0.30Å estimated from the σ_A plot.

The quality of a crystal structure may be assessed by the deviation of the geometry of the model structure from ideal values. Values for deviations from ideality of some geometrical parameters for CI-2 are given in Table 4.3; a more detailed summary of the peptide bond geometry is given in Table 4.4. The mean values of the observed parameters all deviate from the expected values by less than 0.7 times the rms deviation. The deviation of minimum or maximum values from the mean does not exceed 3 times the rms deviation. The peptide bond geometry of the native CI-2 molecule is thus as good as or better than that of the two complex structures.

The distribution of the main-chain torsional angles ϕ and ψ for the native CI-2 is shown in Figure 4.6, a Ramachandran plot (Ramakrishnan & Ramachandran, 1965). This

Table 4.4

Summary of Peptide Bond Geometry

	Bond type	No.	Mean	RMS deviation	Minimum deviation	Maximum deviation	Expected' value
CI-2 Native	N-C^{α}(Å)	65	1.467	0.010	1.454	1.484	1.474
	C^{α}-C(Å)	65	1.527	0.008	1.510	1.543	1.524
	C=O(Å)	65	1.243	0.006	1.229	1.262	1.243
	C-N(Å)	64	1.319	0.007	1.301	1.331	1.323
	τ(°)	65	111.2	3.7	102.9	117.8	109.65
	ω(°)	64	179.9	2.7	-174.9	173.8	180.00

'Expected values from Arnott & Dover (1967).

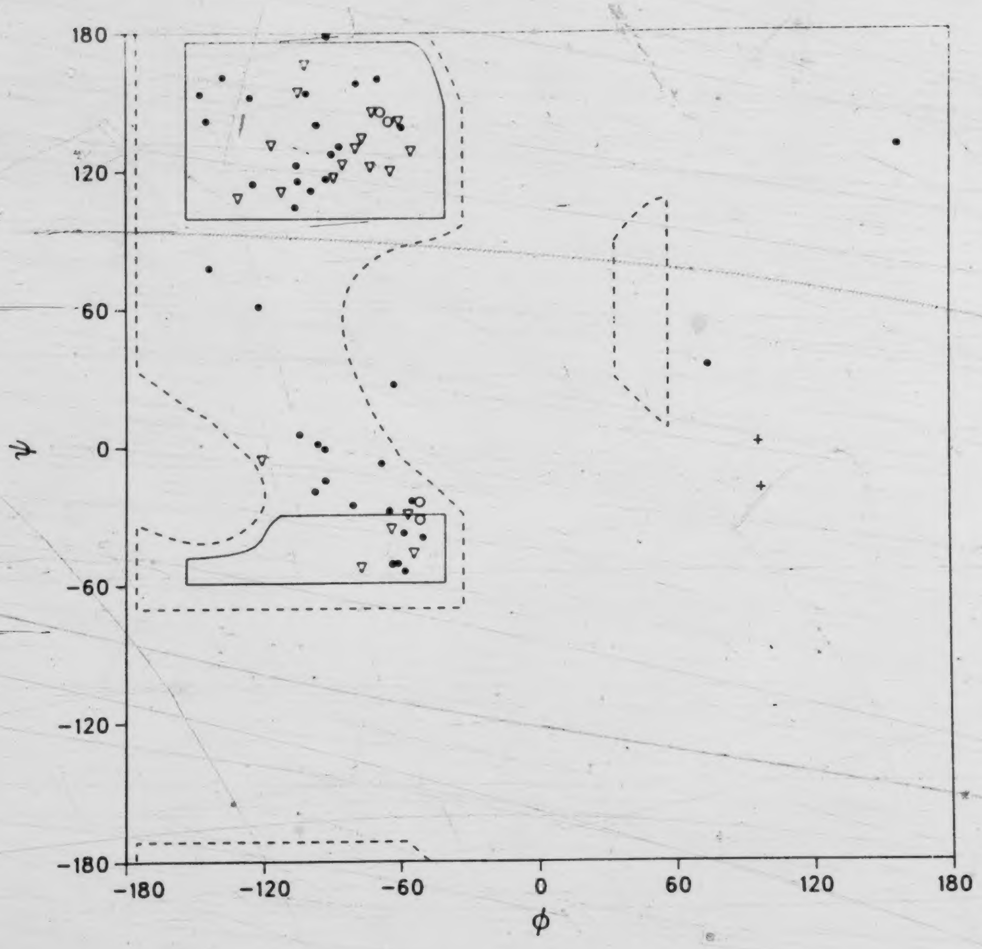


Figure 4.6. ϕ-ψ Plot for CI-2. The solid lines enclose the fully allowed conformational regions for τ (N-C^{α}-C) of 110°; the dotted lines enclose the more permissive regions of smaller acceptable van der Waals' contacts for τ of 115° (Ramakrishnan & Ramachandran, 1965). The symbols denote prolines (o), β-branched amino acids (Δ), glycines (+), and all other amino acids (•). Residues outside the allowed conformational regions and their ϕ-ψ angles are Leu20I (157,130), Met59I (-63,27), and Asp74I (74,35).

distribution can be compared to that in Figure 2.10, the Ramachandran plot for CI-2 in the complex with subtilisin Novo. The greatest difference between the two plots is that the native CI-2 has two more non-glycine residues with ϕ-ψ angles outside the allowed regions. Asp74I is outside the allowed regions in both native and complexed inhibitors. In the native CI-2, Leu20I and Met59I are also outside the allowed regions. All 3 residues are in regions of high *B*-factor, and the electron densities of Leu20I and Met59I are weak, so the torsion angles of these residues are not well-defined.

The final electron density map calculated for the native CI-2 structure is of excellent quality in some regions and rather poor quality in others. The regions of poorest quality are those of residues with high atomic *B*-factors. Regions of weak density are described in Table 4.5; to facilitate comparison with the structures of the two complexes, all residues were examined for weak density at a contour level of 0.56e/Å^3. Examples of regions of good and poor density are given in Figure 4.7.

One final indication of the quality of a crystallographic structure is the *R*-factor. The *R*-factor for the refined structure of the native CI-2 molecule, for all data in the resolution range 8.0 to 2.0Å with $I \geq \sigma(I)$, is 0.198. This value was calculated on 4471 reflections, 76% of the unique set of reflections measured. The *R*-factor calculated on the complete unique set is 0.293. The criteria of quality discussed in this section indicate that: 1) the estimated errors in atomic positions are small and comparable to the errors measured for structures refined in a similar resolution range but with significantly lower

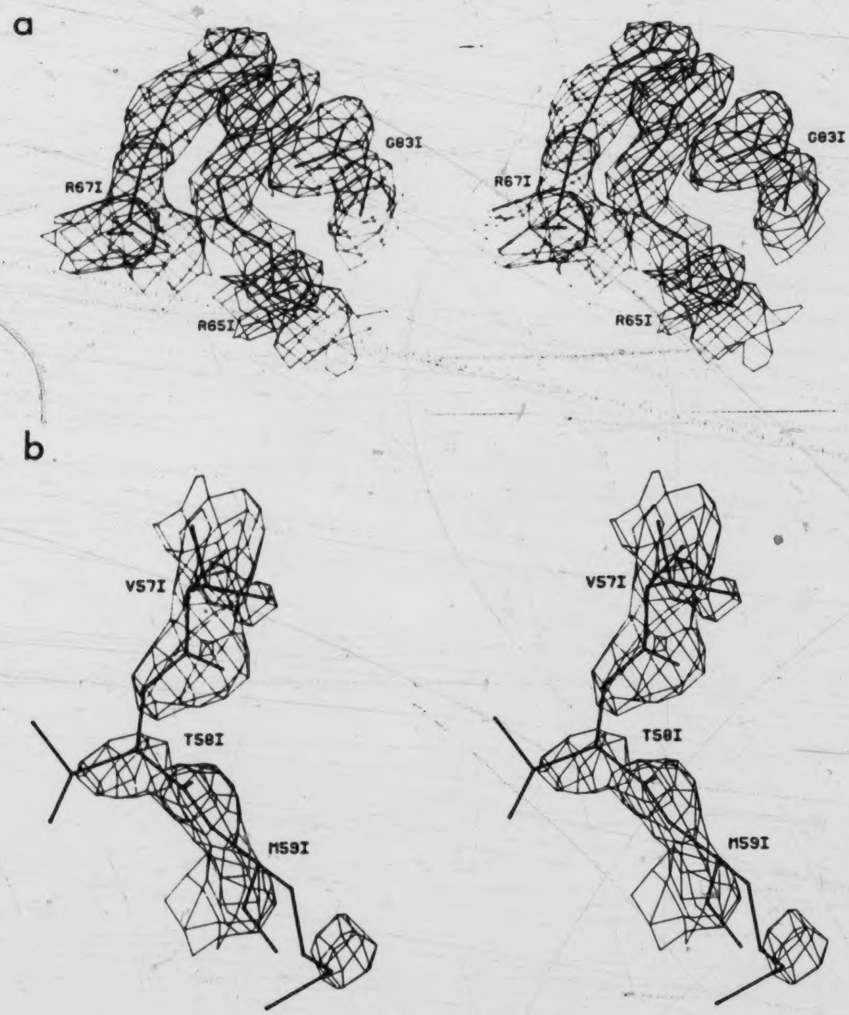

a
G83I
R67I
R65I
G83I
R67I
R65I
b
V57I
T58I
M59I
V57I
T58I
M59I

Figure 4.7. Regions of Good and Poor Electron Density. Both figures are contoured at a level of 0.56e/Å^3. a) Residues Arg65I, Arg67I, and Gly83I, supporting the reactive site loop of the inhibitor. b) Residues Val57I, Thr58I, and Met59I of the reactive site loop. Atoms in this segment of the protein have high *B*-factors.

R-factors; 2) the geometric parameters of the structure do not deviate significantly from ideal values; 3) the indicated shifts in atomic parameters at the final refinement cycle are small, and there are few indications on a difference electron density map that the model requires major adjustments. The structure is of good quality, and the refinement appears to have converged. The relatively high final *R*-factor, indicating only moderately good agreement between observed and calculated structure factors,

Table 4.5

Residues with Poor Electron Density at Cycle 133

Residue	Comment
Asn19I	Density only for C^{α} & carbonyl oxygen.
Leu20I	Density only for C^{β}, C^{γ}, carbonyl carbon & oxygen.
Lys21I	No density for C^{β} & C^{γ}.
Thr22I	No density for $C^{\gamma 2}$.
Glu23I	No density for $O^{\epsilon 2}$.
Lys30I	No density past C^{γ}.
Glu33I	No density for C^{γ} & $O^{\epsilon 1}$. Possible alternate conformations at $\chi_1 \pm 90°$.
Glu34I	No density for C^{γ}.
Ala35I	Weak density for C^{β}.
Lys36I	No density past C^{γ}.
Lys37I	No density past C^{β}.
Ile39I	No density for $C^{\delta 1}$.
Gln41I	No density past C^{β}.
Ile48I	No density for $C^{\delta 1}$.
Ile56I	No density past C^{α}.
Val57I	No density for $C^{\gamma 2}$.
Thr58I	No density past C^{β}, chain break between N & C^{α}.
Met59I	No density for C^{β}, C^{γ}, C^{ϵ}, carbonyl oxygen.
Glu60I	C^{β} out of density.
Arg62I	No density for C^{β}.
Ile63I	No density for $C^{\gamma 1}$, $C^{\delta 1}$, no density between C^{α} & C^{β}.
Lys72I	No density past C^{γ}.
Leu73I	No density for $C^{\gamma 1}$, weak at C^{β}.
Asp74I	Weak density for C^{β}.
Arg81I	No density past C^{γ}.

seems to be due in large part to the high overall *B*-factor of the molecule in the crystal. The high *B*-factor is correlated with the larger percentage of weak electron density observed in the maps and the resulting difficulties in fitting the model to the maps.

4.2 The Structure of the Native CI-2 Inhibitor

4.2.1 Secondary and Tertiary Structures

The structure of the native CI-2 molecule is close to that of CI-2 in the complex with subtilisin Novo in most regions. The two inhibitor structures are compared in detail in chapter 5. Two views of the native inhibitor molecule are shown in Figure 4.8; Table 4.6 lists the secondary structural elements of the native CI-2 molecule and the hydrogen bonds between residues comprising those elements.

The secondary structural elements of the native and complexed CI-2 molecules are almost identical; one hydrogen bond of a β-bridge between the N-terminal and C-terminal strands of the 'pseudo' β-sheet is missing (Val32I to Asp74I), and the hydrogen bonding in the α-helix is more irregular in the native structure. Otherwise, even the main-chain hydrogen bonds that are not part of secondary structural elements are conserved.

The extensive hydrogen-bonding network supporting the reactive site loops in the structures of CI-2 and eglin-c in

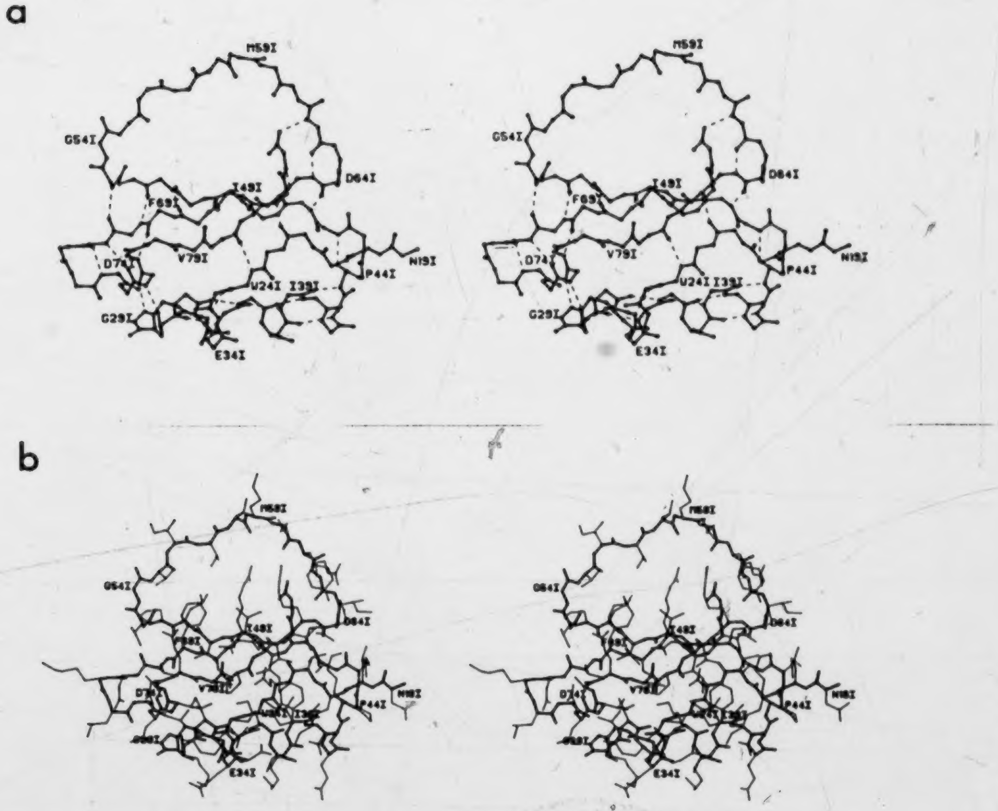

a
M59I
G54I
D64I
I49I
F69I
D74I
V79I
N19I
P44I
W24I
I39I
G29I
E34I
b

Figure 4.8. Views of CI-2. Both views include only residues seen in the electron density map of the native CI-2 molecule, Asn19I to Gly83I. a) The main-chain atoms of CI-2 and their hydrogen-bonding interactions. Hydrogen bonds are drawn with dashed lines. b) All atoms of the CI-2 molecule. The main-chain is drawn with heavy lines and the side-chains with light lines. Every fifth amino acid residue is labelled in both views.

the complexes is partly maintained in the native CI-2 structure. There are some differences in the hydrogen bonds of

Table 4.6

Secondary Structural Elements of CI-2

Element	Residues[1]	N-H···O length (Å)	N-Ĥ-O angle (°)
Parallel β-sheet	Gln47I-Asp64I	2.62	153
	Val66I-Gln47I	2.84	172
	Ile49I-Val66I	2.89	172
	Leu68I-Ile49I	2.54	168
	Leu51I-Leu68I	2.80	165
	Val70I-Leu51I	2.81	165
	Val53I-Val70I	2.99	151
Antiparallel β-bridges	Val82I-Thr22I	2.86	151
	Trp24I-Pro80I	2.91	163
	Asp71I-Asn75I	3.11	158
	Ile76I-Lys30I	2.75	158
	Ala77I-Phe69I	2.83	172
α-helix	Ala35I-Ser31I	2.80	164
	Lys36I-Val32I	2.90	151
	Ile39I-Ala35I	3.09	154
	Leu40I-Lys36I	2.79	163
	Asp42I-Val38I	3.10	158
	Lys43I-Ile39I	2.62	136
Type I turns	Ala46I-Lys43I	2.95	162
	Arg65I-Arg62I	3.11	157
	Asp74I-Asp71I	2.88	157
Type II turn	Lys30I-Leu27I	3.11	161
Type III turn	Leu27I-Trp24I	2.83	151
Unclassified	Gly29I-Ile76I	2.64	159
	Arg62I-Gly83I	2.78	148
	Gly83I-Arg65I	2.96	158

[1] The residue listed first in each pair is the hydrogen bond donor.

the native and complex structures, due mainly to contacts of the native inhibitor with symmetry-related molecules. These can be seen by comparing Figures 2.16, 3.11, and 4.9, and are discussed in chapter 5. The side-chain of Thr58I is turned by -90° in χ_1 in the native structure and makes hydrogen bonds to Glu78I of a symmetry-related molecule. The position of this side-chain is not defined by the electron density map given in Figure 4.7b, but at contour levels of $0.35e/\text{Å}^3$ there is adequate density for the O^γ of Thr58I. The electron density is continuous with that of Glu78I from the neighboring molecule, and the O^γ of Thr58I is the obvious atom to position in the density to form a hydrogen bond. There is no density for the C^γ even at the lower contour level.

The mean geometric parameters of various classes of hydrogen bonds in the native CI-2 structure are given in Table 4.7. All of these values are within two standard

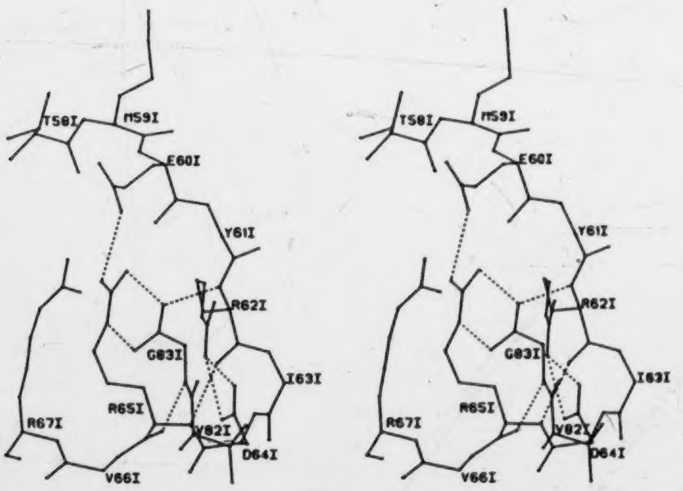

T58I
M59I
E60I
Y61I
R62I
G83I
I63I
R65I
R67I
V82I
D64I
V66I
T58I
M59I
E60I
Y61I
R62I
G83I
I63I
R65I
R67I
V82I
D64I
V66I

Figure 4.9. Hydrogen Bonding of the Reactive Site Loop in Native CI-2. Dashed lines indicate hydrogen bonds. Side-chains of residues not involved in hydrogen bonding have been omitted for clarity.

deviations of the values found for the two complex structures, and most values are within one standard deviation.

4.2.2 Solvent Structure

The final refined model of CI-2 contains 64 solvent molecules, all refined as water. Calculations of the amount of solvent to be expected in the asymmetric unit of these crystals indicate that 25 to 30% of the solvent has been modelled. This is about twice the amount found for the two complex structures, although the same criteria were used in choosing and retaining solvent molecules in the model for all structures (section 2.1.4). Although the native CI-2 molecule has a higher overall *B*-factor than the inhibitor in the complex, most of the difference is due to the active site region (see Fig. 5.7). The remainder of the molecule would be expected to have about the same number of ordered water molecules associated with it. In addition, a much higher percentage of the water molecules associated with the native inhibitor are buried between symmetry-related molecules. The higher percentage of ordered solvent present in the native crystals may be due the greater numbers of these 'trapped' and more ordered waters. For the calculations on protein accessibility and packing in this section, symmetry-related water molecules within 4.0Å of the reference protein molecule were added to the model; there are 26 of these. The complete solvent structure of the native CI-2 molecule, including the 26 symmetry-related waters, is shown in Figure

Table 4.7

Hydrogen-bonding Parameters for CI-2

Type	No.[1]	N···O(Å) (O···O)[2]	H···O(Å)	N-H-O(°)
Main-chain to main-chain				
α-Helix(5→1)	6	2.88(0.19)	1.95(0.15)	157(12)
Antiparallel β-sheet	5	2.89(0.13)	1.94(0.14)	160(8)
Parallel β-sheet	7	2.78(0.16)	1.81(0.16)	164(9)
3_{10} turns(4→1)	5	2.98(0.13)	2.03(0.12)	158(4)
Main-chain to side-chain				
N-H···O	5	2.92(0.19)	1.99(0.14)	159(17)
C=O···H-N	3	2.86(0.18)	1.93(0.15)	154(7)
C=O···H-O	1	2.69(0.00)	-[3]	-
Side-chain to side-chain				
N-H···O	4	2.94(0.15)	1.98(0.14)	164(15)
Main-chain to solvent				
N-H···O	6	2.97(0.17)	2.01(0.18)	165(14)
C=O···O	18	2.89(0.31)	-	-
Side-chain to solvent				
N-H···O	5	3.05(0.14)	2.08(0.15)	165(8)
O(H)···(H)O	8	2.80(0.33)	-	-

[1] Limits for acceptance of a hydrogen bond: donor-acceptor ≤ 3.40Å, hydrogen-acceptor ≤ 2.40Å, donor-hydrogen-acceptor angle ≥ 125°.
[2] Values in parentheses are the sample standard deviations from the mean.
[3] Values are not given for donors with ambiguous hydrogen atom positions.

4.10.

Almost 45% of the solvent molecules within 4.0Å of the surface of CI-2 are buried in the crystal environment i.e. in the presence of other solvent and symmetry-related molecules (less than 10% of their total surface area accessible). Of these 39 buried water molecules, however, only 3 are

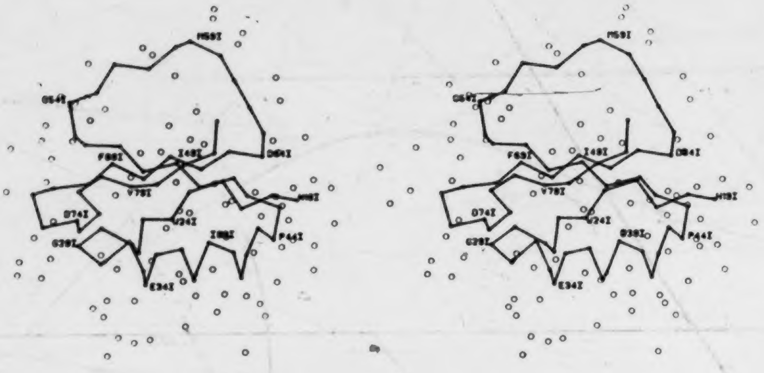


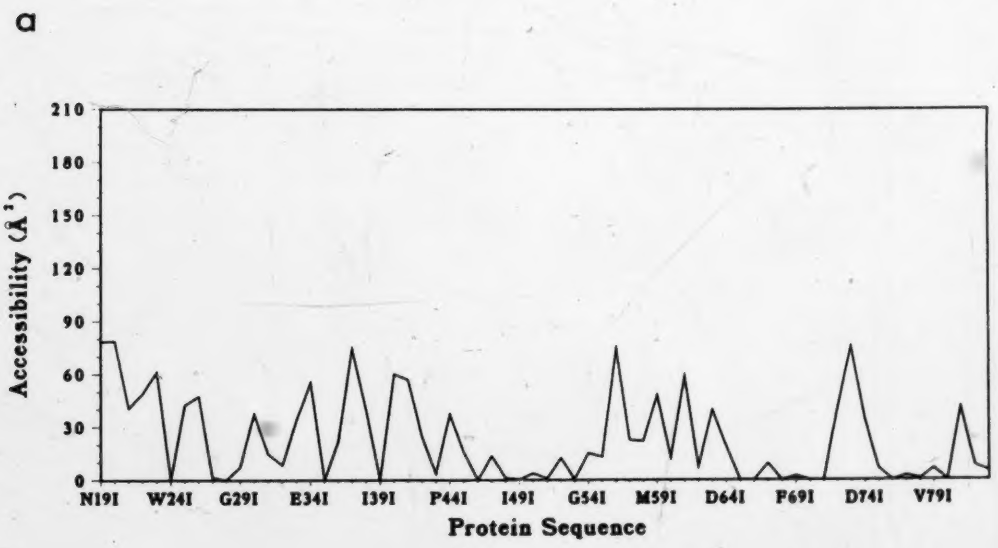
Figure 4.10. Solvent Molecules Associated With CI-2. Water molecules are indicated by open circles and the α-carbon back-bone of CI-2 with thick lines. All solvent molecules in the model of the crystal structure and all symmetry-related solvent molecules within 4.0Å of a protein atom have been included in this figure.

still buried in the absence of other solvent and symmetry-related molecules. These 3 lie between strands of the 'pseudo' β-sheet in CI-2 and provide some of the hydrogen-bonding bridges stabilizing the sheet. The majority of the buried water molecules, 59%, are buried at the interfaces with symmetry-related protein molecules.

Water molecules make a total of 47 hydrogen bonds with polar protein oxygen and nitrogen atoms, and another 32 bonds with other water molecules. The distribution of the hydrogen bonds to protein atoms is close to that found in the larger complex structures; 21% to main-chain nitrogen atoms, 36% to main-chain oxygen atoms, 11% to side-chain nitrogen atoms, and 32% to side-chain oxygen atoms.

4.2.3 Accessibility of Protein Residues

The variation in the accessibility of the protein residues in the native CI-2 molecule is shown in Figure 4.11. There are 6 potentially charged residues with less than 5% of their potential accessible surface area exposed in the molecule in the crystal environment: Lys43I, Arg65I, Arg67I, Asp71I, Glu78I, and Gly83I. Two of these, Lys43I and Gly83I, may still be considered buried when solvent and symmetry-related molecules are removed from the calculation. The side-chain of this lysine is involved in hydrogen-bonding and electrostatic interactions with the side-chain of Glu45I, the carbonyl oxygen atom of Ile63I, and one of



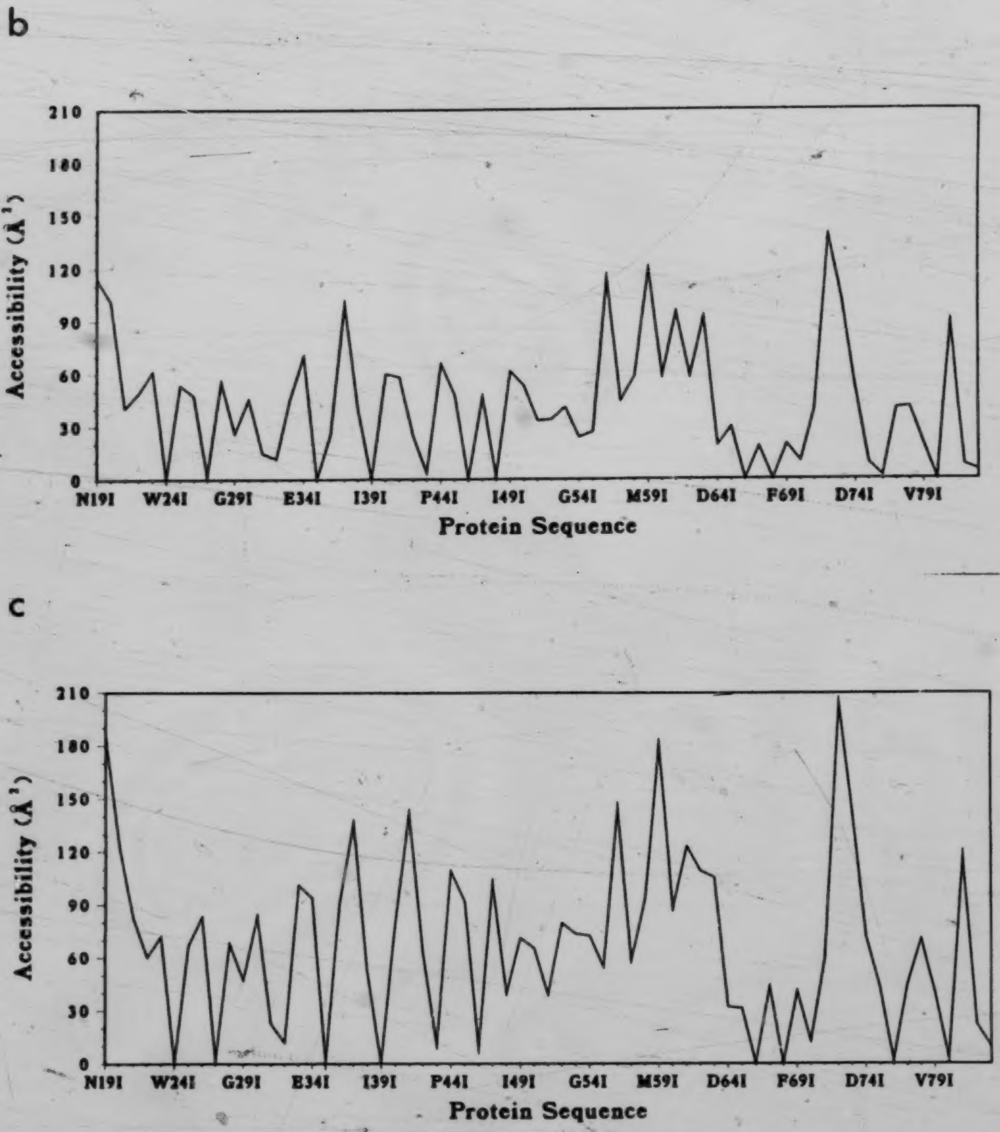


Figure 4.11. Residue Accessibilities in CI-2. Accessibilities are calculated for each atom in the molecule by the method of Lee & Richards (1971). Residue accessibilities are the sum of atomic acessibilities, in Å². a) Accessibility in the presence of all solvent and symmetry-related molecules. b) Accessibility with symmetry-related protein molecules removed but all solvent within 4.0Å of the protein surface retained, including symmetry-related solvent. c) Accessibility in the absence of all solvent and symmetry-related molecules.

the 3 buried water molecules mentioned in the previous section (Fig. 4.8b). Lys43I is not buried in the CI-2 molecule in complex with subtilisin Novo, using the criteria given here. The carboxylate group of Gly83I, the C-terminal residue, forms hydrogen bonds with the amide nitrogen of Arg62I, the N^{ϵ} and $N^{\eta 2}$ of Arg65I, and a buried water molecule (Fig. 4.9).

The reactive site loop of CI-2 is relatively well-buried by symmetry-related molecules in the crystal. The accessibilities of 17 residues in CI-2 from the complex show changes in their residue accessibilities in the presence and absence of the enzyme: Ile49I, Leu51I to Arg62I, Arg67I, Glu78I, Arg81I, and Gly83I. 1000Å^2 of surface area of these residues is buried on complex formation. The change in accessibility of the same residues in the native inhibitor in the presence and absence of symmetry-related molecules is 500Å^2.

Bibliography

Arnott, S. & Dover, S. D. (1967) *J. Mol. Biol. 30*, 209-212.

Barry, C. D., Molnar, C. E., & Rosenberger, F. U. (1976) *Technical Memo No. 229*, Computer Systems Lab, Washington University, St. Louis, MO.

Blow, D. M. & Crick, F. H. C. (1959) *Acta Cryst. 12*, 794.

Bode, W., Epp, O., Huber, R., Laskowski, M. Jr., & Ardelt, W. (1985) *Eur. J. Biochem. 147*, 387-395.

Bragg, W. L. & Perutz, M. F. (1954) *Proc. Roy. Soc. A225*, 315.

Crowther, A. R. (1973) in *The Molecular Replacement Method*, International Science Review *13*, ed. M. G. Rossmann, New York: Gordon & Breach, 173-178.

Cruickshank, D. W. J. (1949) *Acta Cryst. 2*, 65-82.

Cruickshank, D. W. J. (1954) *Acta Cryst. 7*, 519.

Cruickshank, D. W. J. (1967) in *International Tables for X-ray Crystallography* Vol. 2, eds. J. S. Kasper & K. Lonsdale, Birmingham: Kynoch Press, 318-340.

Davies, D. R. & Segal, D. M. (1971) in *Meth. Enzymol.* Vol. XXII, ed. W. B. Jacoby, New York: Academic Press, 266-269.

Hendrickson, W. A. & Konnert, J. H. (1980) in *Biomolecular Structure, Function, Conformation and Evolution* Vol. I, ed. R. Srinivasan, Oxford: Pergamon Press, 43-57.

Jonassen, I. (1980) *Carlsberg Res. Commun. 45*, 47-58.

Kamiya, N., Matsushima, M., & Sugino, H. (1984) *Bull. Chem. Soc. Jpn. 57*, 2075-2081.

Laskowski, M. Jr. & Kato, I. (1980) *Ann. Rev. Biochem. 49*, 593-626.

Lee, B. & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379-400.

Loebermann, H., Tokuoka, R., Deisenhofer, J., & Huber, R. (1984) *J. Mol. Biol. 177*, 531-556.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Cryst. B39*, 480-490.

Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491-497.

Matthews, B. W. (1974) in *The Proteins* Vol. 3, eds. H. Neurath & R. L. Hill, New York: Academic Press, 404-590.

McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, New York: John Wiley & Sons, 96-97.

Mitsui, Y., Satow, Y., Watanabe, Y., & Iitaka, Y. (1979) *J. Mol. Biol. 131*, 697-724.

North, A. C. T., Phillips, D. C., & Mathews, F. S. (1968) *Acta Cryst. A24*, 878-884.

Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I., & Laskowski, M. Jr. (1982) *J. Mol. Biol. 158*, 515-537.

Ramakrishnan, C. & Ramachandran, G. N. (1965) *Biophys. J. 5*, 909-933.

Rao, S. N., Jih, J.-H., & Hartsuck, J. A. (1980) *Acta Cryst. A36*, 878-884.

Read, R. J. (1986) *Acta Cryst.*, in press.

Rossmann, M. G. (1973) *The Molecular Replacement Method*, International Science Review 13, New York: Gordon & Breach.

Seemüller, U., Fritz, H., & Eulitz, M. (1981) in *Methods in Enzymology* Vol. 80, New York: Academic Press, Inc., 804-816.

Sielecki, A. R., James, M. N. G., & Broughton, C. G. (1982) in *Computational Crystallography*, ed. D. Sayre, Oxford: Oxford University Press, 409-419.

Svendsen, I., Boisen, S. & Hejgaard, J. (1982) *Carlsberg Res. Commun. 47*, 45-53.

Svendsen, I., Jonassen, I., Hejgaard, J., & Boisen, S. (1980b) *Carlsberg Res. Commun. 45*, 389-395.

Svendsen, I., Martin, B., & Jonassen, I. (1980a) *Carlsberg Res. Commun. 45*, 79-85.

Thiessen, W. E. & Levy, H. A. (1973) *J. Appl. Cryst. 6*, 309.

5. Discussion and Comparisons of Structures[1]

In this chapter the structures of the subtilisin Novo:CI-2 complex, the subtilisin Carlsberg:eglin-c complex, and the native CI-2 molecule are compared. The primary purpose of these comparisons is the detection of features in the structures that may be relevant to the biological activities of these proteins. Comparisons are also made among inhibitors from other serine proteinase inhibitor families, CI-2, and eglin-c, and among the chymotrypsin-like serine proteinases, subtilisin Novo, and subtilisin Carlsberg.

Most of the comparisons were made after overlapping two molecules by a least-squares superposition of a selected group of common atoms. All of the least-squares superpositions were performed with a program written by W. Bennett. Except for the superpositions of entire molecules, the common atoms overlapped were chosen such that as few assumptions as possible were made about which atoms in a structure should be equivalently positioned.

---

[1] Versions of portions of this chapter have been published [McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242-7246; McPhalen, C. A., Schnebli, H. P., & James, M. N. G. (1985) *FEBS Letters 188*, 55-58].

5.1 Subtilisin Novo and Subtilisin BPN'

The refined structure of subtilisin Novo from the complex with CI-2 can be compared with the structure of the subtilisin BPN' molecule that was used in the molecular replacement solution for the complex. The 275 α-carbon atoms in each molecule can be overlapped with an rms deviation of 0.64Å in atomic positions. Comparison of the unrefined structures of subtilisin Novo (Drenth *et al.*, 1972) and subtilisin BPN' (Wright *et al.*, 1969) gave an rms deviation of 1.14Å in the α-carbon positions, on the order of the errors estimated for the atomic positions in those structures. The positions of 3 α-carbon atoms differ by more than 2.0Å: Gly100, Ser101, and Gly102. The loop containing these residues appears to lie closer to the enzyme active site in the complex, perhaps to form the necessary hydrophobic and hydrogen-bonding interactions for inhibitor binding. The true magnitude of this change is difficult to assess, however, since the electron density for the loop in the subtilisin BPN' structure is poor (Alden *et al.*, 1971) and the coordinate errors must be large. Three other residues close to the loop have α-carbon positions differing by more than 1.0Å: Gly97, Asp98, and Gly128. These changes are probably connected with the changes in the position of the loop upon inhibitor binding.

An additional 10 residues in the proteins have α-carbon positions differing by more than 1.0Å. None of the 10 are close to the active site, and all are part of exterior loops with relatively high *B*-factors. These changes may be due to relatively poor determination of the positions of the residues in the loops, rather than any real difference in the atomic positions. The changes could also be due to differences in packing contacts in the crystals. Detailed comparisons between the structures of the native and complexed enzymes are difficult to make because the coordinates of the native structure available from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) are only partially refined to an *R*-factor of 0.44 and may still contain coordinate errors on the order of 1.0Å (Alden *et al.*, 1971).

Alden *et al.* (1971) have included 17 well-ordered water molecules in their model of subtilisin BPN'. If the coordinates of the waters are transformed by the matrix relating the coordinates of the α-carbon atoms of the two enzymes, 6 of the BPN' water molecules are within 1.0Å of subtilisin Novo water molecules. All 6 are buried water molecules in the subtilisin Novo structure and are some of the best-defined water molecules in the complex. Three of the 6 are part of the chain of water molecules between the side-chains of His67, His226, and the bulk solvent (Fig. 2.23). Two are associated with Asp32, and the last is buried in the bottom of the P_1 specificity pocket.

5.2 Subtilisin Carlsberg and Subtilisin Novo

5.2.1 Global Comparison

The refined structure of subtilisin Carlsberg in complex with eglin-c will be compared with the refined structure of subtilisin Novo in complex with CI-2. A least-squares superposition of the 274 structurally equivalent C^{α} atoms in the two enzymes (omitting Thr55 of subtilisin Novo) gives an rms deviation of 0.53Å, a relatively small value that indicates substantial similarity between the two structures (Fig. 5.1). Sixteen C^{α}'s differ in position by more than 1.0Å, two of these by just over 2.0Å (53 and 159); residues 16 to 20, 52 to 56 (deletion at residue 55 in subtilisin Carlsberg), 131 and 132, 159 to 162,

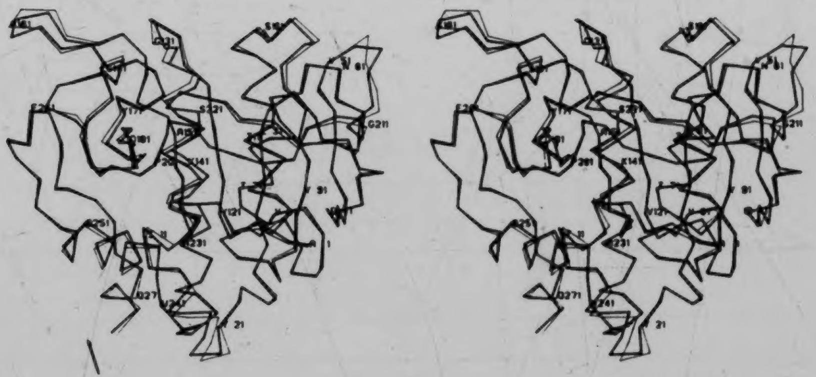


Figure 5.1. Superposition of Subtilisins Carlsberg and Novo. The 274 structurally equivalent α-carbon atoms of the two enzymes were overlapped. Subtilisin Novo is drawn with thin lines, subtilisin Carlsberg with thick lines. Every tenth residue is labelled for subtilisin Novo.

and 211.

Residues 16 to 20 are part of helix B, on the surface of the enzyme and far from the active site; this region has contacts with a symmetry-related molecule in the subtilisin Novo crystals, but not in the Carlsberg crystals. The shift in position of helix B is associated with smaller shifts, on the order of 0.6Å, in helices G and I which pack against B. No associated shift is seen in helix F, which also packs against B; thus the shift in B is not propagated to the active site. Residues 52 to 56 are part of the loop containing Thr55 in subtilisin Novo, which is deleted in subtilisin Carlsberg. This is also an exterior loop with relatively high *B*-factors. Residues 131 and 132 are part of an exterior loop that changes position in concert with a shift in the position of the side chain of Tyr104. The tyrosine side-chain may be shifted as a result of differences in contacts with the bound inhibitors, particularly the P_4 side-chain. Residues 159 to 162 are part of poorly ordered external loops in both enzymes (Fig. 3.6b) and the errors in the atomic coordinates of such regions are higher (Figs. 2.7 and 3.3), thus the differences in the α-carbon positions may be less significant. Residue 211 is the threonine preceded by a *cis*-peptide bond in subtilisin Carlsberg; it is a glycine in subtilisin Novo, with ϕ-ψ angles of 70°, 48°, quite normal angles for a glycine.

There are 82 amino acid differences between subtilisins Carlsberg and Novo (Fig. 2.1). The effects of these sequence changes on the structures of the enzymes were examined by overlapping the two molecules with the matrix relating the positions of the equivalent α-carbon atoms, and looking at the overlapped molecules on the graphics system. About 40% of the differences are in residues that lie on the surface of the molecule and protrude into the solvent; they have only a small effect on the course of the protein backbone. Another 15% are in surface crevices or the protein interior, but the changes are conservative and the side-chains occupy essentially the equivalent space in each enzyme. Several changes are less conservative in terms of the space occupied by the side-chains, but hydrogen-bonding atoms overlap closely e.g. the substitutions (Carlsberg to Novo) Thr33→Ser33, Thr78→Ser78, Ser183→Asn183, and Ser242→Thr242.

In several cases, a single amino acid change from Carlsberg to Novo is very non-conservative but is compensated by another change close by in space; the net effect of the changes is that two side-chains together occupy the equivalent space in each enzyme. Examples of this are Tyr57 and Asn58 in Carlsberg occupying the equivalent space of Asn56 and Phe58 in Novo, Asn97 and Ser99 to Gly97 and Asp99, Tyr143 and Ala243 to Val143 and Asn243, Glu197 and Ser251 to Asp197 and Gln251, and Leu241 and Arg249 to Trp241 and Ser249. A somewhat more elaborate compensatory substitution is that of Gly63, Tyr209, and Leu217 in Carlsberg for Ser63, Leu209 and Tyr217 in Novo (see Fig. 5.2).

In 6 instances, substitutions of longer side-chains allow the formation of new salt bridges or hydrogen bonds: Gln10 (Carlsberg) with Asp181, Gln36 (Carlsberg) with Asn212, Gln59 (Novo) with Asn61, Lys141 (Carlsberg) with Glu112, Ser158 (Carlsberg) with Arg186, and Asn163 (Carlsberg) with Glu195. The position of the main-chain of the external loop from Ser158 to Thr162 in Carlsberg is more than 1.0Å from its position in Novo; this difference is associated with the formation of the new hydrogen bond between Ser158 and Arg186 in Carlsberg.

Other changes in the position of the main-chain are associated with the substitution of a larger side-chain at the position of a buried residue: Ile165→Val165 (Carlsberg to Novo) with the shift of Glu195, Val205→Ile205 with the shift of Pro201, Ala254→Thr254 with the shift of its own main-chain atoms. The course of the main-chain also differs about residue 211, the site of the second *cis*-peptide in Carlsberg. The C-terminus of subtilisin Carlsberg lies closer to the side-chain of Arg249 (substituted for Ser249 in subtilisin Novo). Finally, the changes of Ala52→Pro52 (Novo to Carlsberg) and Ser103→Gln103 allow the formation of a surface crevice in Novo that fits around the side-chain of Trp106 (Gly106 in Carlsberg).

5.2.2 Active Site Regions

The active site regions of subtilisin Carlsberg and subtilisin Novo can be compared after the superposition of the 274 equivalent α-carbon atoms of the two enzymes. The residues of the catalytic triad, His64, Ser221, and Asp32, have essentially identical conformations at this resolution; most equivalent atoms of these 3 residues are less than 0.1 to 0.2Å apart, and none are further than 0.3Å. Other residues with differences in atomic positions in the range of 0.1 to 0.3Å are Thr33, Ser125, Gly128, Thr220, and Met222. Residues Asn62, Ala129, and Asn155 overlap to within 0.5Å for equivalent atoms. The loop of residues Leu96 to Ser105 has shifted from its position in subtilisin Novo to lie closer to the inhibitor in subtilisin Carlsberg. The shifts for equivalent atoms in this loop are in the range of 0.5 to 0.9Å. In the crystal structure of the native subtilisin Novo, this loop was described as very mobile and flexible (Drenth *et al.*, 1972; Wright *et al.*, 1969). The different positions of the loop in the two enzyme:inhibitor complexes may also indicate some degree of induced fit in the formation of the complex. In addition to the general change in the loop, the side-chain of Tyr104 is rotated 90° in χ_1 and out toward the solvent, changing the shape of the S_4 binding pocket on the enzyme.

The active sites of subtilisin Novo and subtilisin Carlsberg may also be superposed by first overlapping functional groups in the region, rather than the global

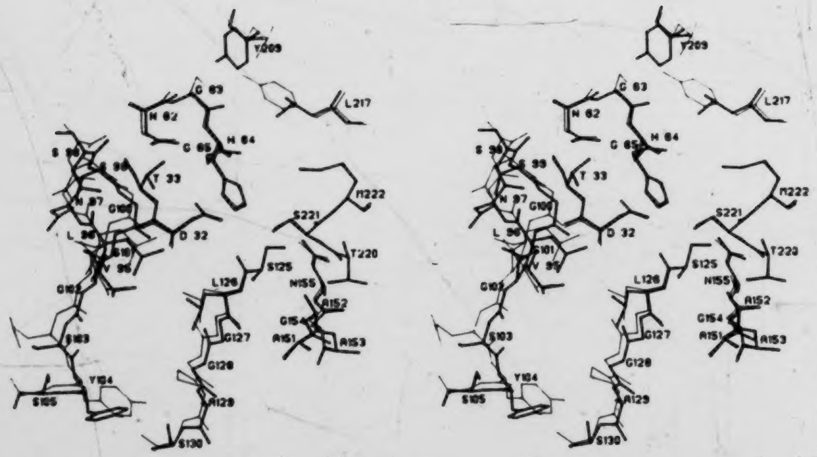

Y209
G 63
L217
N 62
H 64
G 65
S 98
S 99
T 33
M222
N 97
G100
S221
L 96
D 32
S101
T220
V 95
S125
L126
N155
G102
A152
G154
A151
A153
S103
G127
G128
Y104
S105
A129
S130

Figure 5.2. Superposition of Active Sites of Subtilisins. Functional groups in the active site regions of subtilisin Carlsberg and subtilisin Novo were overlapped as described in the text. Subtilisin Novo is drawn with thin lines, subtilisin Carlsberg with thick lines. The residues of subtilisin Carlsberg are labelled.

superposition of the α-carbon atoms. The superpositions of two sets of functional groups were examined. First, all atoms of the catalytic histidine and serine residues, plus the main-chain atoms of the P_1 and P_1' residues were overlapped. The inhibitor atoms were included to compare the geometry of the active sites in relation to the position of bound inhibitor/substrate. The rms deviation in position of the 26 superposed atoms is 0.22Å. The active site regions of the two enzymes overlapped with the matrix relating the functional groups are shown in Figure 5.2.

The second superposition of functional groups overlapped all main-chain and C^{β} atoms (except for glycine residues) of the following segments in Carlsberg and Novo:

Leu31 to Gly34, Asn62 to Thr71, Met124 to Gly128, Ala151 to Asn155, and Leu217 to His225. These segments include sequences flanking the catalytic triad, and residues forming the walls of the S_1 specificity pocket and the oxyanion hole. The rms deviation in position of these 161 superposed atoms is 0.19Å, very close to the 0.22Å found for the first, much more limited, functional superposition.

The results of the two superpositions of functional groups indicate that the active site regions of subtilisin Novo and subtilisin Carlsberg are much more similar than implied by the global superposition of α-carbon atoms. Even residues not directly involved in substrate binding and catalysis have almost identical conformations, and the rms deviation between the two sets of atoms in both functional superpositions is close to the estimated errors in atomic positions for the subtilisin Novo structure.

After the functional superpositions, the main-chain atoms of the loop from Leu96 to Ser105 still differ by 0.6 to 1.3Å. Considering how closely the rest of the active site regions overlap, this is a significant difference that points strongly to an adjustment of this part of the active site to the conformation of different inhibitors.

Overall, the few changes in the active site region may have some effect on the specificity or strength of substrate/inhibitor binding, but the general similarity of the catalytic apparatus and binding sites indicates that there should be little difference in catalytic activity between Carlsberg and Novo. The observed differences in the ratios of esterase to proteinase activity (Barel & Glazer, 1968) must thus be due to more subtle conformational differences than those discussed here, or to other factors such as differences in the electrostatic environment of the two active sites.

5.2.3 Solvent Structure

A number of solvent molecules are conserved between the subtilisin Novo:CI-2 and subtilisin Carlsberg:eglin-c complexes. A solvent molecule was considered to be conserved if, after the two enzymes were overlapped by the global superposition of α-carbon atoms described above, the solvent molecule had a counterpart in the second structure less than 1.0Å away. There were 45 water molecules that met this criterion. Some of the conserved water molecules have an obvious structural role, such as providing hydrogen-bonding interactions to a buried polar residue. Conserved waters in this category include the channel of 6 waters from the side-chains of His67, Thr71, and His226 to the enzyme surface, and buried waters adjacent to the side-chains of Asp32, Ser190, and Thr253 (Fig. 2.23). The water trapped at the bottom of the S_1 specificity pocket is conserved, as is one water molecule buried between strands of the β-sheet in the inhibitors. The second water molecule conserved between the inhibitors is one bound to the C-terminal carboxyl group of Gly83I.

Thirty-three water molecules are conserved in crevices on the surfaces of the enzymes. Six of the conserved waters on the surface lie in a crevice close to Leu75 and the loop forming the binding site of Ca^{2+} ion 1. An acetone molecule from the crystallization medium was observed in this area in the structure of the native subtilisin Novo (Drenth *et al.*, 1972; Hol, 1971).

The two Ca^{2+} ions bound to subtilisin Carlsberg are conserved in subtilisin Novo; the K^+ ion is replaced by a water molecule with an occupancy of 1.0 and a *B*-factor of 12Å². The two ordered water molecules coordinating the K^+ ion in subtilisin Carlsberg are not conserved in subtilisin Novo. The difference in the positions of the Ca^{2+} ions after the global superposition of α-carbon atoms of the two enzymes is 0.36Å for Ca^{2+} ion 1 and 0.91Å for Ca^{2+} ion 2. As was seen for the active site region, however, a local superposition of functionally important atoms can reveal much stronger structural homologies.

Ca^{2+} site 1 from the two enzymes was superposed by a least-squares fit of equivalent atoms from the protein residues acting as ligands to the Ca^{2+}; included were all atoms of Gln2, Asp41, Leu75, Asn77, main-chain and C^{β} atoms of Thr79 (Ile79 in Novo), and all atoms of Val81. The rms deviation of these atoms after the superposition was 0.14Å, less than the estimated rms coordinate error for the subtilisin Novo complex. The difference in the positions of the Ca^{2+} ions after the superposition was 0.12Å, significantly

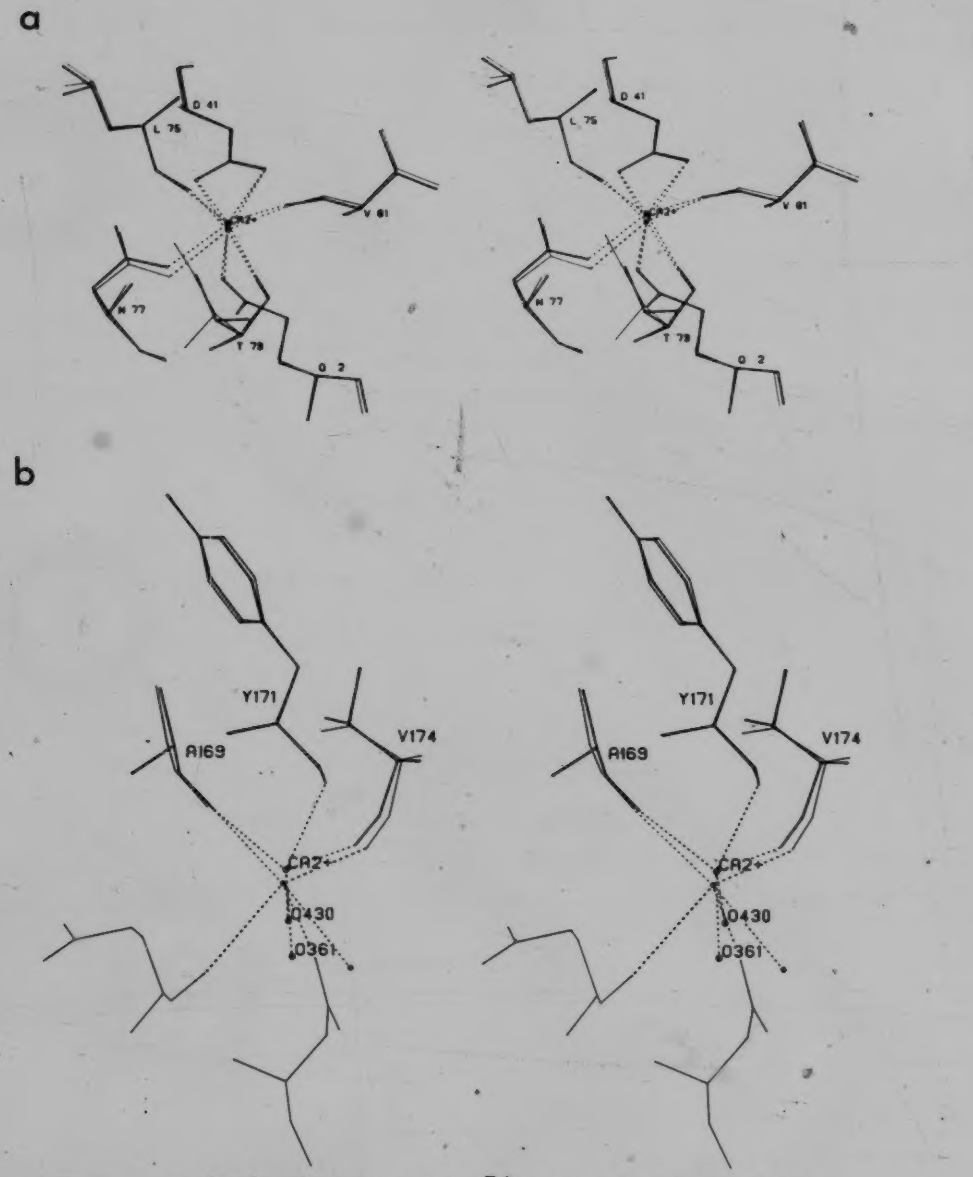


Figure 5.3. Superposition of Ca^{2+} Ion Binding Sites. Functional groups of the Ca^{2+} ion binding sites of subtilisin Carlsberg and subtilisin Novo were overlapped as described in the text. Subtilisin Novo is drawn with thin lines, subtilisin Carlsberg with thick lines. The residues of subtilisin Carlsberg are labelled. a) Ca^{2+} ion binding site 1. b) Ca^{2+} ion binding site 2. The unlabelled residues of subtilisin Novo are Glu195 and Asp197.

smaller than the value obtained from the global superposition. The superposed binding sites from the two enzymes are shown in Figure 5.3a.

A similar functional superposition was performed for the binding site of Ca^{2+} ion 2; all atoms except C^{β} of Ala169 (Gly169 in Novo), and all atoms of Tyr171 and Val173 were included. The results of this superposition are shown in Figure 5.3b. The rms deviation of the superposed atoms is 0.16Å, close to the value for Ca^{2+} site 1. The Ca^{2+} ions from the two enzymes are 0.35Å apart, again a significantly smaller value than that obtained from the global superposition. The coordinating ligand O361 in Carlsberg is 0.81Å from the $O^{\delta 2}$ of Asp197 in Novo, but both the other two ligands in Novo are more than 2.0Å away from any possible corresponding ligand in Carlsberg. The geometry of Ca^{2+} binding site 2 is thus much less well conserved than that of site 1. This may be related to the relative importance of the two ions to enzyme stability; Ca^{2+} ion 1 is proposed to be more important to the integrity of the active site and the hydrophobic core of the enzyme than Ca^{2+} ion 2 (section 2.2.1).

5.3 Subtilisin Carlsberg and α-Chymotrypsin

Previous comparisons of the active site regions of the native subtilisin Novo and α-chymotrypsin concluded that the geometry of the catalytic triad was virtually identical in the two enzymes (Wright, 1972; Kraut *et al.*, 1972), although the remainder of the molecules are different. A similar comparison has been made between subtilisin Carlsberg from the complex with eglin-c and a refined α-chymotrypsin structure from the complex with OMTKY3 (*R*-factor = 0.168 for data in the resolution range from 8.0 to 1.8Å (Read *et al*., 1984)). The α-chymotrypsin from an enzyme:inhibitor complex was chosen for the comparison to minimize possible differences in the conformation of catalytic residues as a result of substrate/inhibitor binding. The structure of subtilisin Carlsberg was used for the comparison because it has been refined at higher resolution than the subtilisin Novo complex.

The active sites of the enzymes were superposed by a least-squares fit of the side-chains of the catalytic histidine and serine residues, C^α of the histidine, C^α and N of the serine, plus the main-chain atoms of the P_1 and P_1' residues of the respective inhibitors. This set of atoms was chosen by visual inspection on the MMS-X graphics system after a rough manual superposition. Atoms clearly not equivalent were excluded (e.g. the main-chain of the catalytic histidine goes in opposite directions in the two enzymes). Assumptions about the equivalence of residues (e.g. the catalytic Asp) were avoided as much as possible. The conformations of the inhibitors' P_1 and P_1' main-chain atoms are virtually identical. The inhibitor atoms were included in order to compare the geometry of the active sites in relation to the position of bound substrate/inhibitor. The rms deviation of the 19 superposed atoms is 0.46Å after the least-squares minimization. The distances between the equivalent atoms included in the superposition are given in Table 5.1. The two active site regions overlapped after the superposition are shown in Figure 5.4.

The side-chains of the two active site histidines and serines, His64 and Ser221 in Carlsberg, His57 and Ser195 in α-chymotrypsin, superpose well, indicating that their relative orientation is similar in both enzymes. The active site aspartates, Asp32 and Asp102, are quite different in position and orientation. In α-chymotrypsin, the carboxyl

Table 5.1

Distances Between Equivalent Atoms, Subtilisin Carlsberg and α-Chymotrypsin

Carlsberg atom	Chymotrypsin atom	Distance (Å)
His64 C^{α}	His57 C^{α}	0.30
His64 C^{β}	His57 C^{β}	0.87
His64 C^{γ}	His57 C^{γ}	0.33
His64 $N^{\delta 1}$	His57 $N^{\delta 1}$	0.45
His64 $C^{\delta 2}$	His57 $C^{\delta 2}$	0.50
His64 $C^{\epsilon 1}$	His57 $C^{\epsilon 1}$	0.57
His64 $N^{\epsilon 2}$	His57 $N^{\epsilon 2}$	0.54
Ser221 NH	Ser195 NH	0.12
Ser221 C^{α}	Ser195 C^{α}	0.31
Ser221 C^{β}	Ser195 C^{β}	0.52
Ser221 O^{γ}	Ser195 O^{γ}	0.42
Leu59I NH	Leu18I NH	0.27
Leu59I C^{α}	Leu18I C^{α}	0.29
Leu59I C	Leu18I C	0.24
Leu59I O	Leu18I O	0.46
Asp60I NH	Glu19I NH	0.12
Asp60I C^{α}	Glu19I C^{α}	0.33
Asp60I C	Glu19I C	0.75
Asp60I O	Glu19I O	0.54

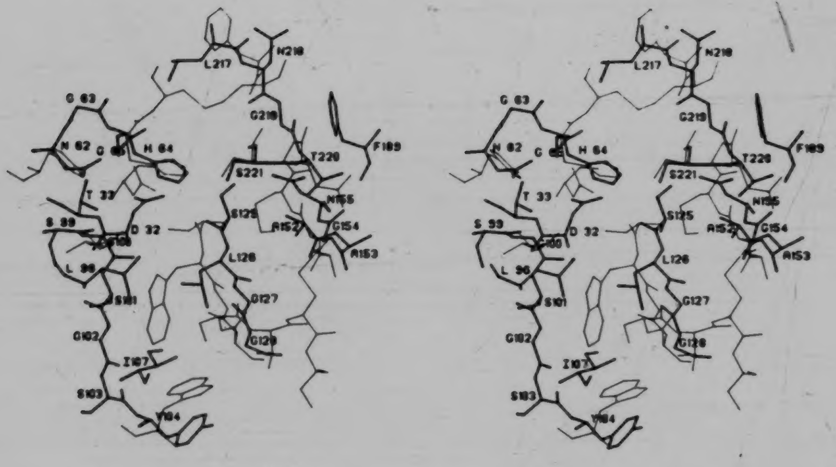

N218
L217
G 63
G219
F189
N 62
H 64
T220
S221
T 33
N155
S125
S 99
D 32
A152
G154
A153
L126
L 96
S101
G127
G102
I107
G128
S103
Y104

Figure 5.4. Superposition of Subtilisin Carlsberg and α-Chymotrypsin Active Sites. The active site regions of the two enzymes were overlapped as described in the text. α-Chymotrypsin is drawn with thin lines, subtilisin Carlsberg with thick lines. The residues of subtilisin Carlsberg are labelled.

group of Asp102 is at an angle of 30° to the imidazole ring of His57; there is one hydrogen bond from $N^{\delta 1}$ of the histidine to $O^{\delta 1}$ of the aspartic acid, and $O^{\delta 2}$ points behind the histidine toward the protein interior. The O^{γ} of Ser214 provides another hydrogen bond to $O^{\delta 1}$, and $O^{\delta 2}$ receives hydrogen bonds from the main-chain nitrogens of Ala56 and His57. In Carlsberg, Asp32 is somewhat less buried; its carboxyl group is also at an angle of 30° to the plane of the imidazole ring of His64, but $O^{\delta 2}$ points in front of the histidine toward the protein exterior. The angle between the two carboxyl groups in the overlapped proteins is thus 60°. Both carboxyl oxygen atoms of Asp32 form hydrogen bonds to His64 $N^{\delta 1}$, but the strongest interaction is with $O^{\delta 2}$. The second hydrogen-bonding partner for $O^{\delta 2}$ is the main-chain nitrogen of Thr33, and $O^{\delta 1}$ interacts with a buried water molecule. The geometry of these interactions in subtilisins Novo, Carlsberg, and α-chymotrypsin is summarized in Table 5.2.

Apart from the catalytic triad, there are three other close superpositions of atoms. The carbonyl oxygens of Ser125 (Carlsberg) and Ser214 (α-chymotrypsin), which are believed to be important in the formation of the tetrahedral enzyme:substrate intermediate (Robertus *et al.*, 1972), are 0.9Å apart and both oriented toward the main-chain nitrogen of the inhibitors' P_1 residue; the oxygen - nitrogen distance is 3.64Å in the subtilisin Carlsberg:eglin-c complex and 3.33Å in the α-chymotrypsin:OMTKY3 complex. The S^{δ} of Met222 is 0.8Å from the S^{γ} of Cys42, above and in the plane of the His-Ser catalytic interaction. Substitution of amino acids without sulphur atoms in their side-chains at position 222 in subtilisin results in specific activities for the mutant enzymes of 0 to 50% of the native activity (Estell *et al.*, 1985). The role of the sulphur atom in serine proteinase catalysis is unknown; it may play an electrostatic role, but this has not been investigated to date. The carbonyl oxygens of Asn218 and Phe41 are 0.9Å apart; each forms a hydrogen bond with the main-chain NH of the P_2 residue in their respective inhibitor. The hydrogen bonds and short contact distances seen in the active site regions of 4

Table 5.2

Hydrogen Bonds of the Catalytic Triad

Bond[1]	N···O(Å)	H···O(Å)	N-$\hat{H}$-O(°)
His64 $N^{\epsilon 2}$			
Novo - Ser221 O^{γ}	2.73	1.79	154
Carlsberg - Ser221 O^{γ}	2.68	1.70	165
Chymotrypsin[2] - Ser195 O^{γ}	2.51	1.83	123
His64 $N^{\delta 1}$			
Novo - Asp32 $O^{\delta 1}$	3.11	2.20	150
Novo - Asp32 $O^{\delta 2}$	2.79	1.83	159
Carlsberg - Asp32 $O^{\delta 1}$	3.14	2.27	145
Carlsberg - Asp32 $O^{\delta 2}$	2.68	1.71	162
Chymotrypsin - Asp102 $O^{\delta 1}$	2.76	1.78	168
Asp32 $O^{\delta 1}$			
Novo - O352 O	2.72	-	-
Carlsberg - O344 O	2.64	-	-
Chymotrypsin - Ser214 O^{γ}	2.61	1.61	113
Asp32 $O^{\delta 2}$			
Novo - Ser33 NH	2.72	1.79	153
Carlsberg - Thr33 NH	2.69	1.81	145
Chymotrypsin - Ala56 NH	2.76	2.29	107
Chymotrypsin - His57 NH	2.85	1.93	151

[1] His64 and Asp32 are equivalent to His57 and Asp102 in α-chymotrypsin.
[2] The values for α-chymotrypsin are from its structure in complex with OMTKY3 (Read *et al.*, 1984).

serine proteinase:inhibitor complexes are given in Table 5.3.

Other components of the active site regions are shifted by more than 1.0Å, or are constructed from different foldings of the polypeptide chain. The strand Gly100 to Gly102 in Carlsberg, which forms another antiparallel β-interaction with the inhibitor, has no counterpart in α-chymotrypsin and is distant from any residues of that enzyme. Residues Ser125 to Gly127 in Carlsberg and Ser214

Table 5.3

Contacts (Å) in the Active Sites of Serine Proteinase Complexes

Site		Carlsberg: Eglin-c	Novo: CI-2	SGPB: OMTKY3[1]	Trypsin: PTI[2]
P_6	I CO - Tyr104 OH	2.5	-	-	-
P_4	I NH - Gly102 CO	-	2.9	-	-
	I CO - Gly102 NH	3.4	3.0	-	-
P_3	I NH - Gly127 CO[1]	2.9	3.0	2.9	3.3
	I CO - Gly127 NH	2.9	3.0	3.0	-
P_2	I NH - Gly100 CO	2.9	3.1	-	-
P_1	I NH - Ser125 CO	3.6	3.4	3.6	3.4
	I CO - Ser221 NH	3.2	3.0	3.1	3.1
	I C - Ser221 O^{γ}	2.8	2.7	2.7	2.6
	I CO - Asn155 $N^{\delta 2}$	2.8	2.7	2.6	2.9
P_2'	I NH - Asn218 CO	2.8	2.8	2.8	-
	I CO - O370 O	2.6	2.7	3.2	3.0

[1] The equivalent residues in SGPB and trypsin are: Ser125→Ser214, Gly127→Gly216, Asn155→Gly193 NH, Asn218→Phe41, and Ser221→Ser195. Data from Read *et al.*, 1983.

[2] Data calculated from coordinates obtained from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).

to Gly216 in α-chymotrypsin both form a short stretch of antiparallel β-sheet with inhibitor residues P_1 to P_3, but this strand in α-chymotrypsin is shifted more than 1.0Å away from the P_1 side-chain. An accompanying shift of the residues in the other wall of the specificity pocket in α-chymotrypsin makes the width of the pockets in the two enzymes comparable. Carlsberg lacks residues corresponding to Ser218 to Ser221 in α-chymotrypsin, which loop up to form a wall at the end of the P_1 specificity pocket; the distal end of the pocket is much more open in Carlsberg, and the subtilisins thus have a more broad substrate specificity.

The side-chain of Gly166 of the subtilisins, at the bottom of the P_1 specificity pocket, has been mutated to each of the other 19 amino acids to study its effect on substrate specificity (Wells *et al*., 1985). Specificity can be altered significantly, depending on the size and charge of the amino acid replacing the glycine.

The carbonyl oxygen atom of the P_1 residue points into the oxyanion hole in both complexes and forms a hydrogen bond to the NH of the catalytic serine; the NH's of the two serines are 0.1Å apart in the overlapped active sites. The other components of the oxyanion hole are the $N^{\delta 2}$ of Asn155 in Carlsberg and the NH of Gly193 in α-chymotrypsin, which are 1.5Å apart. Although the two nitrogens are distant, hydrogen atoms placed at ideal positions on the nitrogens would be only about 1.0Å apart. The side-chains of asparagine residues are not normally considered to be as rigid conformationally as the nitrogen of a peptide bond. Asn155, however, is involved in additional hydrogen bonds to the NH and O^{γ} of Thr220 and is relatively fixed in position (Fig. 3.12). The *B*-factors of the $O^{\delta 1}$ (10Å²) and $N^{\delta 2}$ (9Å²) of the Asn155 side-chain reflect the lack of flexibility; the average *B*-factors for these atoms in the 18 other asparagines in the enzyme are 23Å² for $O^{\delta 1}$ and 19Å² for $N^{\delta 2}$.

The S_2 specificity site in subtilisin Carlsberg is a hydrophobic pocket bounded by the plane of the imidazole ring of His64, the side-chain of Leu96, and the main-chain of Gly100. Because α-chymotrypsin lacks the loop containing

Gly100, its S_2 specificity pocket is much more open. This is not true of all chymotrypsin-like enzymes, however; in the complex of SGPB:OMTKY3, an additional wall for the S_2 pocket is provided by the side-chain of Tyr171 (Read *et al.*, 1983; James *et al.*, 1980) and the shape of the pocket is more like that of subtilisin Carlsberg.

In the Carlsberg complex, the P_4 side-chain of the inhibitor is buried in a deep pocket bounded by Tyr104, Ile107, Leu126, Leu96, Gly127 and Gly128. A smaller cleft is formed in α-chymotrypsin by Trp215, Trp172 and Gly216. The cleft in subtilisin Novo is also smaller because the side-chain of Tyr104 has rotated inward and away from the solvent in subtilisin Novo relative to its position in subtilisin Carlsberg (Fig. 5.2). There is a hydrophobic stacking of the P_2' side-chain with Phe189 in Carlsberg; no comparable residue is present in α-chymotrypsin. Finally, both enzymes have comparable hydrophobic patches, possibly to bind to P_3' residues; this is composed of Leu217 in Carlsberg, and Phe41 plus Gly59 in α-chymotrypsin.

A second functional superposition of the active sites of subtilisin Carlsberg and α-chymotrypsin was performed for comparative purposes. The atoms overlapped were all involved in hydrogen-bonding interactions in the active site: the carbonyl oxygen of Asn217 (Phe41 in α-chymotrypsin), the $N^{\delta 1}$ and $N^{\epsilon 2}$ of His64 (His57), the $O^{\delta 2}$ of Asp32 ($O^{\delta 1}$ of Asp102), the carbonyl oxygen of Ser125 (Ser214), and the amide nitrogen and O^{γ} of Ser221 (Ser195).

This superposition was performed twice, with and without the main-chain atoms of the inhibitor P_1 and P_1' residues. The rms deviations for these two superpositions were 0.60Å and 0.63Å, respectively. The matrices relating the two enzyme active sites were very similar for all 3 functional superpositions, so the above discussion comparing the active sites is not highly dependent on the choice of atoms for the superposition.

5.4 CI-2 and Eglin-c

The structures of eglin-c and CI-2 from the complexes with the subtilisins were overlapped by a least-squares superposition of structurally equivalent α-carbon atoms, Lys22I to Pro72I and Thr73I to Gly83I (Gly72AI is an insertion in eglin-c relative to CI-2 (Fig. 2.2)). The rms difference in position for all these C^{α}'s is 1.68Å. Iteration of the least-squares process and gradual removal of the most widely separated pairs of atoms from the calculation leads to an rms deviation of 0.73Å for 39 out of the 62 α-carbon pairs. Residues included in the calculation of this final rms deviation are Lys22I (eglin-c sequence) to Lys30I, Glu45I to Glu53I, Leu59I to Asn69I, and Val76I to Gly83I. The regions of best agreement between the two inhibitors are thus the 4 strands of β-sheet and the second part of the reactive site loop (Fig. 5.5).

Although eglin-c and CI-2 have only moderate sequence homology (Fig. 2.2) and show some relatively large

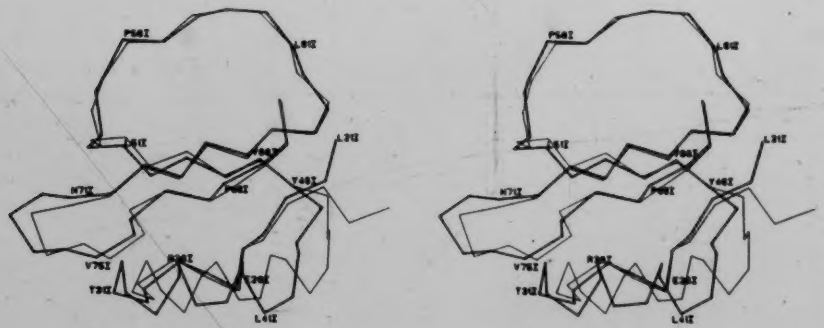


Figure 5.5. Superposition of Eglin-c and CI-2. The α-carbon atoms of the two inhibitors were overlapped as described in the text. CI-2 is drawn with thin lines, eglin-c with thick lines. The residues of eglin-c are labelled.

conformational differences, major structural features are very well conserved. These features include the hydrophobic core, the central β-sheet, and the hydrogen-bonding network supporting the reactive site loop. The residues of the α-helix show deviations of 1.9 to 5.2Å between α-carbon atoms. In eglin-c, this helix has a number of packing contacts with symmetry-related molecules that do not occur in CI-2. These contacts may be partly responsible for the shifted helix position (also see below).

A second segment of eglin-c that deviates in position from CI-2 is Gly54I to Thr58I, the first part of the reactive site loop (2.1 to 4.1Å difference in α-carbon atoms). The enzyme residues interacting with this portion of the reactive site loop, Gly100 to Ser105, have also shifted between the active sites of subtilisin Carlsberg and subtilisin Novo. The third segment differing between the two inhibitors is the exterior loop which contains the inserted residue Gly72AI in eglin-c.

The reactive site loops of eglin-c and CI-2 were overlapped by a least-squares superposition of the main-chain plus C^{β} atoms of the P_1 and P_1' residues, in order to compare the environments of their reactive site bonds. The rms deviation of the 10 pairs of atoms in this overlap was 0.11Å. A second overlap was performed to compare the entire reactive site loops; the atoms overlapped were the

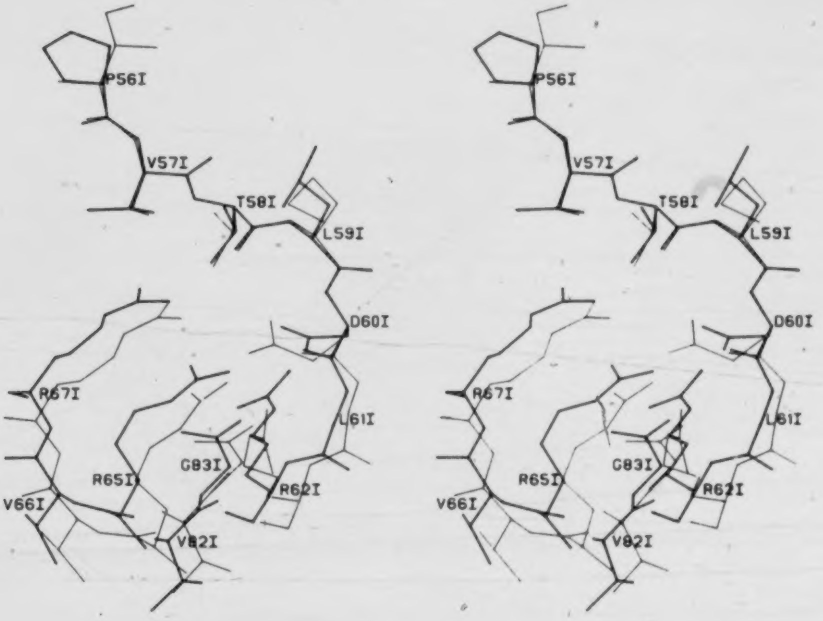

P56I
V57I
T58I
L59I
D60I
R67I
L61I
R65I
G83I
R62I
V66I
V82I
P56I
V57I
T58I
L59I
D60I
R67I
L61I
R65I
G83I
R62I
V66I
V82I

Figure 5.6. Superposition of Eglin-c and CI-2 Reactive Site Loops. The reactive site loops of the two inhibitors were overlapped as described in the text. CI-2 is drawn with thin lines, eglin-c with thick lines. The residues of eglin-c are labelled.

main-chain plus C^{β} atoms of the P_3 to P_1' residues and the main-chain atoms of the P_4 residues (P_4 of eglin-c is a proline). The rms deviation of the atoms in the second comparison was 0.19Å.

The superposed reactive site loops and the surrounding residues that interact with them are shown in Figure 5.6. The results of the second overlap were used for this figure. In this view the very close overlap of the reactive site loops can be seen clearly, including the P_4 to P_1' residues. This is in contrast to the differences seen in this region when the whole inhibitors are overlapped, thus the body of the inhibitor must have shifted relative to the reactive site loop between the two complexes.

There is one major change between the two reactive site loops, the substitution of Asp60I in eglin-c for Glu60I in CI-2. The P_1' side-chain is extensively involved in hydrogen bonding that presumably stabilizes the conformation of the loop and the reactive site bond. The hydrogen-bonding networks of the two inhibitors are shown in Figures 2.16 and 3.11. The $O^{\delta 2}$ of Asp60I and the $O^{\epsilon 2}$ of Glu60I are 0.6Å apart in the superposed reactive site loops; this is relatively close, since one side-chain is one carbon shorter than the other. Part of the compensation for the shorter P_1' side-chain comes from a difference in χ_1 of 30° and a difference in orientation of the two carboxyl groups. A second part of the compensation comes from a shift of Arg62I, Arg65I, Arg67I, and Gly83I up toward Asp60I relative to the positions of these residues in CI-2; the shift can be seen in the positions of the main-chain atoms of the residues in Figure 5.6. The main-chain C^{α}'s of the arginines in the two inhibitors are over 1.0Å apart.

A third part of the compensation comes from the presence of an additional water molecule adjacent to Asp60I $O^{\delta 2}$. This water molecule (O372) provides hydrogen-bonding bridges between Asp60I, Thr58I, and Arg65I. The number of hydrogen bonds in the network and the atoms involved in the bonds are thus highly conserved between the two inhibitors. This is an indication of the importance of the network to the stability of the reactive site loop.

The helix in eglin-c is shifted by an angle of 18° and translated by 0.8Å relative to the helix in CI-2 when the α-carbon atoms of the inhibitors are overlapped. Most of the change in helix position relative to the β-sheet and reactive site loop appears to be due to changes in the packing of the hydrophobic core of eglin-c. The core of eglin-c is formed almost entirely from the same residues as that of CI-2, but almost 75% of these positions contain amino acid substitutions. The core of CI-2 contains 1 aromatic residue and 15 aliphatics; eglin-c contains 7 aromatics and 10 aliphatics.

The reasons for the change in conformation with these substitutions are not obvious from a simple examination of the overlapped helices on the graphics system. The residues of CI-2 were therefore 'mutated' to the corresponding residues in the core of eglin-c (program of R. Read). The 'mutated' residues retain the positions of the main-chain atoms and the appropriate side-chain torsional angles of the original residues. The 'mutated' CI-2 molecule was examined on the graphics for bad contacts of side-chains that might be relieved by a change in conformation. Three bad contacts were found, all of which could be relieved by a shift of the eglin-c helix in the direction of the observed shift. The bad contacts were from the substitution of Tyr46I (eglin-c) for Ala46I (CI-2), Phe50I for Val50I, and Phe39I for Ile39I (Fig. 2.15, Fig. 3.10). Tyr46I was seriously entangled with Tyr43I in the 'mutated' CI-2, Phe50I bumped into Val32I and Arg36I, and Phe39I was in close contact with Phe24I. The shift in the helix of eglin-c also allows a more favorable stacking of the side-chain rings of Tyr38I with Phe39I, Phe24I, and His42I.

5.5 Inhibitor Interactions with Serine Proteinases

A number of standard interactions of inhibitors or substrate analogs bound to serine proteinases have emerged from comparisons of X-ray structures of a variety of complexes (Read & James, 1986, and references therein). The comparisons indicate that the reactive site loops of the inhibitors studied all have similar conformations, and the conformation is presumed to be similar to that of a productive binding mode of a peptide substrate. To a first approximation, the formation of the enzyme:inhibitor complex is by the 'lock and key' mechanism, with little adjustment of the conformation of either partner.

The standard interactions that are seen in all serine proteinase:inhibitor complexes are: specificity pockets for P_4, P_2, and P_1 side-chains, a β-sheet interaction from P_3 to residue 127 in the subtilisins (216 in the chymptrypsin-like enzymes), a long hydrogen bond from the main-chain NH of the P_1 residue to residue 125 CO (214), a close approach ($\simeq$ 2.7Å) of Ser221 O^γ (195) to the carbonyl carbon atom of P_1, and a hydrogen bond from the P_2' main-chain NH to the CO of residue 218 (41). Complexes of inhibitors with subtilisins exhibit an additional set of β-interactions with the loop of residues 100 to 104 as seen in the structures of SSI (Hirono *et al.*, 1984), CI-2, and eglin-c with subtilisins.

An examination of the details of the interactions of inhibitors with serine proteinases shows considerable diversity within and beyond the standard interactions. The hydrogen bonds between enzymes and inhibitors are somewhat variable (Table 5.3). For example, one half of the β-interaction between P_3 and residue 216 is missing in the trypsin:PTI complex, as is the hydrogen bond from the main-chain NH of P_2' to residue 41 (Read *et al.*, 1983). The hydrogen-bonding interactions in subtilisin:inhibitor complexes are summarized in Table 5.4, and show many differences between complexes. In some cases, the differences are due to changes in the sequences of the reactive site loops,

Table 5.4

Hydrogen Bonds of Subtilisin:Inhibitor Complexes

Site	Bond	Eglin-c	CI-2	SSI[1]	Hypothetical Substrate[2]
P_6	I CO - Tyr104 NH			✓	
	I CO - Tyr104 OH	✓			
P_5	I CO - Ser130 NH				✓
P_4	I NH - Gly102 CO		✓	✓	
	I CO - Gly102 NH	✓	✓		
P_3	I NH - Gly127 CO	✓	✓	✓	✓
	I CO - Gly127 NH	✓	✓	✓	✓
P_2	I NH - Gly100 CO	✓	✓		
P_1	I NH - Ser125 CO	long	long	✓	
	I CO - Ser221 NH	✓	✓		
	I CO - Thr220 NH			✓	
	I CO - Asn155 $N^{\delta2}$	✓	✓	✓	
P_2'	I NH - Asn218 CO	✓	✓		✓
	I CO - Asn218 NH				✓

[1] Hirono *et al.* (1984).
[2] Robertus *et al.* (1972).

particularly the presence of prolines at P_4 in eglin-c and P_2 in SSI.

Other differences occur in interactions with the loop from Gly100 to Tyr104. As was demonstrated in the comparison of the active sites of subtilisin Novo and subtilisin Carlsberg, this loop has shifted between the two complexes and the side-chain of Tyr104 is in a different position and orientation. In order to form the hydrogen bonds between the loop and the inhibitor shown for the subtilisin:SSI complex, the conformation of the loop would have to be different again in that complex. These data imply that there is a small but significant induced fit between the subtilisins and their inhibitors.

One possible source of the differing loop conformations is the size and shape of the P_4 side-chain. The differences between the P_4 residues of eglin-c and CI-2 can be seen in Figure 5.6. The size and shape of the S_4 specificity pocket is also different in the cognate enzymes, and the loop from Gly100 to Tyr104 forms one wall of the pocket.

5.6 CI-2 Native and CI-2 in Complex

The native CI-2 structure and CI-2 from the complex with subtilisin Novo can be overlapped by a superposition of equivalent α-carbon atoms, Leu20I to Gly83I. The rms deviation in atomic positions is 0.47Å for the α-carbon atoms. The largest deviations in position, between 0.5 and

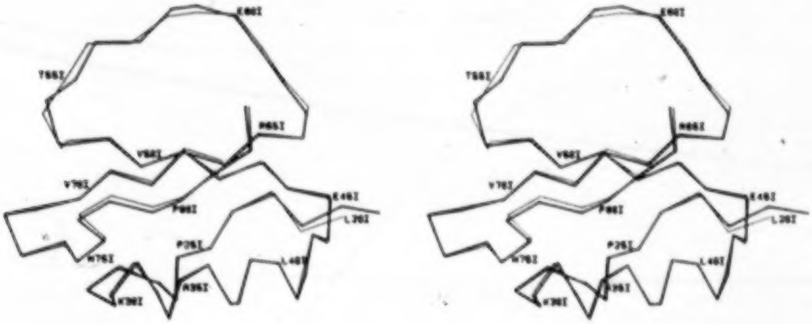


Figure 5.7. Superposition of CI-2 Molecules. The α-carbon atoms of the two inhibitors were overlapped as described in the text. CI-2 from the complex is drawn with thin lines, CI-2 native with thick lines.

1.4Å, occur in the reactive site loop. The body of the inhibitor changes very little between the complex and native structures (Fig. 5.7). Changes in the reactive site loop are difficult to assess; the high *B*-factors in this region in the native structure imply larger coordinate errors, on the order of half the observed rms deviation.

Some of the solvent structure of the CI-2 molecule in the complex with subtilisin Novo is conserved in the native molecule. If the two molecules are overlapped by the matrix relating the α-carbon atoms, there are 7 water molecules in equivalent positions, within 1.0Å of each other. Four of these are the water molecules that form the hydrogen-bonding bridges in the β-sheet of the complexed inhibitor (Fig. 2.17). The other 3 lie close to the C-terminus of the single α-helix.

The difference in the mean *B*-factor of main-chain atoms for each residue is plotted in Figure 5.8. The mean *B*-factor for all main-chain atoms in the native CI-2 is 19; the mean for CI-2 from the complex is 14. In most regions the native CI-2 molecule has *B*-factors only marginally above or below those of the molecule from the complex. Some of the exterior loops in the native structure are better ordered than in the complex structure, probably because they are involved in contacts with symmetry-related molecules in the native CI-2 crystals. The reactive site loop of the native inhibitor is obviously the major region of disorder relative to the inhibitor from the complex, and the major

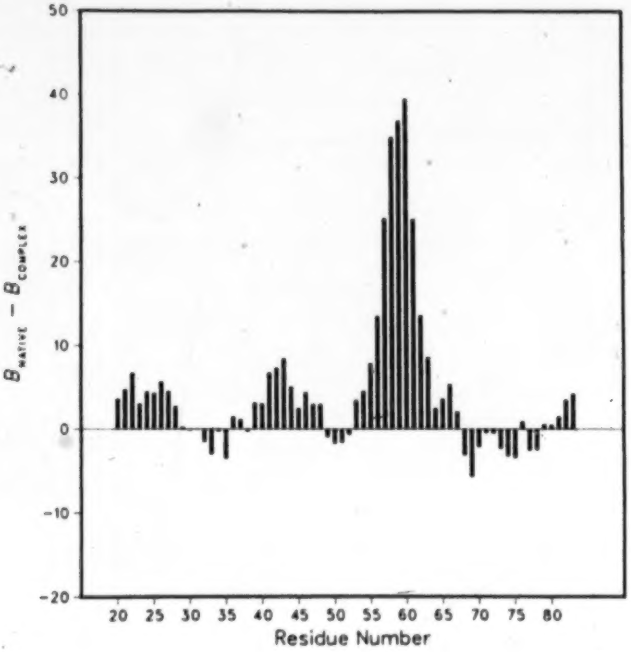


Figure 5.8. Difference *B*-Factor Plot, CI-2 Native vs CI-2 in Complex. The mean *B*-factor of the main-chain atoms was calculated for each residue in native CI-2 and CI-2 from the complex with subtilisin Novo. The difference in mean *B*-factor is plotted versus residue number. No value is given for Asn19I; no density is seen for this residue in the CI-2 structure from the complex.

contributor to the difference in overall *B*-factor.

The differences in position of the reactive site loop do not disappear if a more local superposition is performed. As in the comparison of eglin-c and CI-2 from the complex, the main-chain plus C^{β} atoms of the P_4 to P_1' residues of the loops were overlapped; the rms deviation for this superposition was 0.51Å, more than twice the difference between the eglin-c and CI-2 reactive site loops. Figure 5.9 shows the result of the local superposition. The deviation in position of main-chain atoms is spread

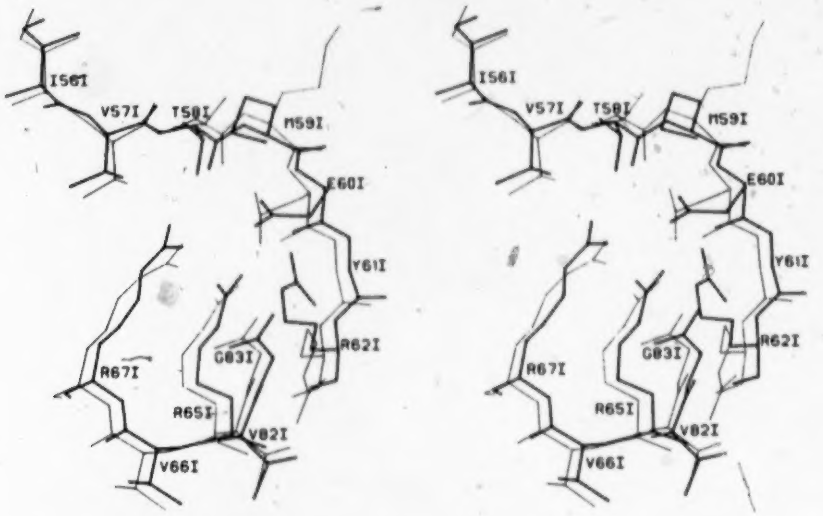


Figure 5.9. Superposition of CI-2 Reactive Site Loops. The reactive site loops of the two inhibitors were overlapped as described in the text. CI-2 from the complex is drawn with thick lines, CI-2 native with thin lines.

throughout the loop region.

The positions of several side-chains in the reactive site loop are also changed between the two inhibitor structures. Thr58I has changed by -90° in χ_1; $O^{\gamma 1}$ in the native inhibitor points away from Glu60I and forms a hydrogen bond with the side-chain of Glu78I of a neighboring molecule. Glu60I in the native inhibitor no longer bends back toward its own main-chain NH, but has a more extended conformation. Arg62I has rotated to form a salt bridge with Asp64I in the native inhibitor, rather than the hydrogen bonds with Glu60I seen in the complex. Met59I is in an extended conformation in the native inhibitor; the less extended conformation in the complex is presumably due to limiting contacts with the S_1 specificity pocket.

The changes in side-chain conformations in the reactive site loop are associated with the loss of several hydrogen bonds from the stabilizing network in the native inhibitor, primarily those formed by Thr58I and Glu60I in the complex. Although it is difficult to assign cause and effect in such matters, the lack of the hydrogen bonds in the native inhibitor may allow freer movement of the reactive site loop, resulting in the high *B*-factors and disorder observed in this region. The arginines supporting the loop retain most of their hydrogen bonds between complex and native, and are well-ordered in both structures (Fig. 4.7).

The relatively high *B*-factors for the reactive site loop of native inhibitors compared to the inhibitors in a complex have been observed previously for inhibitors of the Kazal family (Read & James, 1986). This indication of flexibility in the reactive site loop has been proposed as the mechanism by which inhibitors act on a wider range of enzymes; while the reactive site loop is stabilized by various hydrogen-bonding and covalent interactions to a conformation complementary to an enzyme active site, it it still able to adapt to more than one enzyme active site (Read & James, 1986). The increase in flexibility of an inhibitor reactive site loop results in a less tight binding to cognate enzymes; the entropic cost of the loss of extra internal degrees of freedom in the inhibitor reduces the available binding energy. In evolutionary terms, however, the increase in broadness of inhibitor action may be worth the decrease in tightness of binding.

A change in the relative orientation of the reactive site loop to the body of the inhibitor in complexes with different enzymes has been observed for OMTKY3 in complexes with SGPB and α-chymotrypsin; the change has been described as a hinge-like movement (Read *et al.*, 1984). A similar change can be seen between the complexes of subtilisin Novo:CI-2 and subtilisin Carlsberg:eglin-c. The conformations of the reactive site loops are almost identical (see above) in the two inhibitors, and the majority of the enzyme active sites overlap closely. If the two complexes are overlapped by the matrix relating the two enzymes, however, there is a noticeable twist of one inhibitor relative to the other (Figure 5.10).

Some of the difference in orientation of the two inhibitors is due to the changes in their hydrophobic cores and a shift in the positions of their α-helices. The angle between the two helices in Figure 5.10, however, is 45°, compared to 18° if only the inhibitors are overlapped: at least half of the difference in orientation of the inhibitors in the two complexes is due to other factors. If eglin-c is placed in the active site of subtilisin Novo, using the matrix relating the two enzyme active sites, there is only one contact less than 2.0Å between enzyme and inhibitor. This contact is between the OH of Tyr104 and Gly54I; Tyr104 changes position in subtilisin Carlsberg relative to subtilisin Novo. The lack of bad contacts in a subtilisin

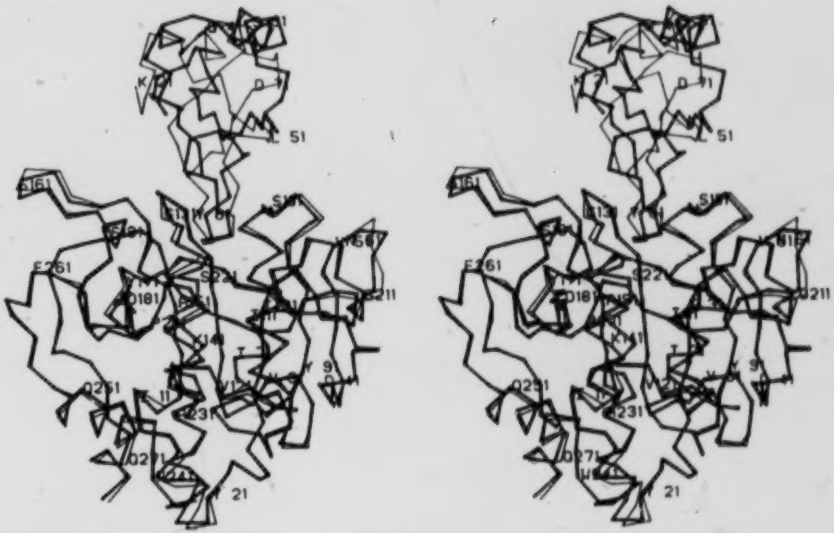


Figure 5.10. Superposition of Enzyme:Inhibitor Complexes. The complexes were overlapped with the matrix relating the two enzymes. The subtilisin Carlsberg:eglin-c complex is drawn with thick lines, subtilisin Novo:CI-2 with thin lines. Residues of subtilisin Carlsberg:eglin-c are labelled.

Novo:eglin-c model complex indicates that the change in inhibitor orientation is not due to changes in enzyme:inhibitor interactions. There are few bad contacts between eglin-c in the model complex with subtilisin Novo and symmetry-related molecules in the subtilisin Novo crystal environment, thus packing contacts should only contribute a small amount to the change in orientation. Flexibility of the reactive site loop region appears to allow the inhibitors to change orientation and still maintain good interactions with the enzyme active site.

The flexibility of the reactive site loop relative to the body of these inhibitors may be due to Arg65I and Arg67I. They act as spacers supporting the reactive site loop, in a role analogous to that of Asn33I of OMTKY3 (Fujinaga *et al.*, 1982; Read *et al.*, 1983); the spacers allow the loop to protrude from the inhibitor surface and interact easily with the enzyme active site. Arginine side-chains are also relatively flexible and would allow the reactive site loop some freedom to adapt to the active sites of different enzymes by changing its position relative to the inhibitor body. The arginines thus have a dual role; they provide flexibility to broaden the specificity of the inhibitor, and they form an integral part of the hydrogen bonding network that stabilizes the conformation of the residues of the reactive site loop to be complementary to the enzyme active site.

5.7 Why Are Eglin-c and CI-2 Good Inhibitors?

According to the standard mechanism of action for protein inhibitors of serine proteinases put forward by Laskowski and Kato (1980), a good inhibitor binds very tightly to the enzyme (low K_M) and is hydrolyzed very slowly (low k_{cat}). The association rate constants for eglin-c with elastase and cathepsin G are on the order of 10^7 M^{-1} sec^{-1} and the equilibrium association constants are 10^{11} M^{-1} (Schnebli *et al.*, 1985). The rapid and tight association of enzyme and inhibitor is promoted by the reactive site region of the inhibitor being held fairly rigidly in a conformation complementary to the enzyme active site.

Both eglin-c and CI-2 seem to conform to the part of the inhibitory mechanism requiring tight binding to the enzyme. The complementarity between their reactive site loops and the enzyme active site is excellent, as seen in previous chapters. The reactive site loop of the inhibitors is stabilized by extensive hydrogen bonding; if the reactive site residues were simply a free peptide in solution, they would presumably be much more labile in conformation than even the reactive site loop of the native inhibitor. Some of the lack of disorder in the reactive site loop of a bound inhibitor thus comes from the stabilizing hydrogen bonds, and the rest from interactions with the enzyme on complex formation. The conformations of the loops of CI-2 and eglin-c are highly similar, as are their interactions with the enzyme active site. In the complex, a number of interactions favor tight binding of the inhibitor to the enzyme. These include the hydrogen bonding in the two sets of β-sheet interactions with subtilisin, as well as extensive hydrophobic interactions of inhibitor side-chains with deep specificity pockets on the enzyme.

The second required property of a good inhibitor is a slow rate of hydrolysis. The K_{hyd} for inhibitors is usually close to 1 near neutral pH (Finkenstadt *et al.*, 1974; Laskowski & Kato, 1980). Values of K_{hyd} are not available for CI-2 and eglin-c, but the similarity of their interactions in enzyme complexes to the interactions of other families of inhibitors indicates that they may share a mechanism. The slow rate of hydrolysis of protein inhibitors of serine proteinases may be due to activation energy barriers at many points in the reaction. Some of the barriers may be due to structural properties of the inhibitors alone; others may be due to interactions between the inhibitors and enzymes.

One activation energy barrier has been proposed for the formation of the tetrahedral intermediate from enzyme and uncleaved inhibitor (Fujinaga *et al.*, 1982; Read *et al.*, 1983). The presence of this type of barrier is in agreement with early observations that modified inhibitors (ones with a cleaved reactive site bond, but with the two halves still joined by disulphide bonds) bind more slowly to α-chymotrypsin than uncleaved inhibitors by a factor of about 10^5 (Quast *et al.*, 1978). The experimental data implied that once an inhibitor was cleaved it would lose its function, thus a good inhibitor would be especially resistant to cleavage. Some of the structural features of OMTKY3 that would hinder the formation of a tetrahedral intermediate on the pathway to cleavage are the hydrogen bonds from the side-chain of Asn33I to the carbonyl oxygen atoms of P_2 and P_1', from the NH of the P_1' residue to its own side-chain, and from the side-chain of P_2 to the side-chain of P_1' (Fujinaga *et al.*, 1982). Hydrogen bonds or electrostatic interactions analogous to these are seen in CI-2 and eglin-c, involving Thr58I, the P_1' residue, Arg65I, and Arg67I (Fig. 2.16 and 3.11). An inhibitor closely related to CI-2, CI-1, has a phenylalanine residue in place of Arg67I (Svendsen *et al.*, 1982). CI-1 is cleaved and dissociated from serine proteinases more rapidly than CI-2, indicating the importance of the arginine residues to inhibition (I. Svendsen, personal communication).

Recent reports have shown that for subtilisin Novo and subtilisin Carlsberg, the ratio of the association rate constants of the enzymes with uncleaved or modified OMTKY3 are close to 1, in contrast to the ratio of 10^6 found with α-chymotrypsin (Ardelt & Laskowski, 1985). Barriers to cleavage of the inhibitor thus may also be important after formation of the acyl enzyme but before release of the cleaved inhibitor from the enzyme.

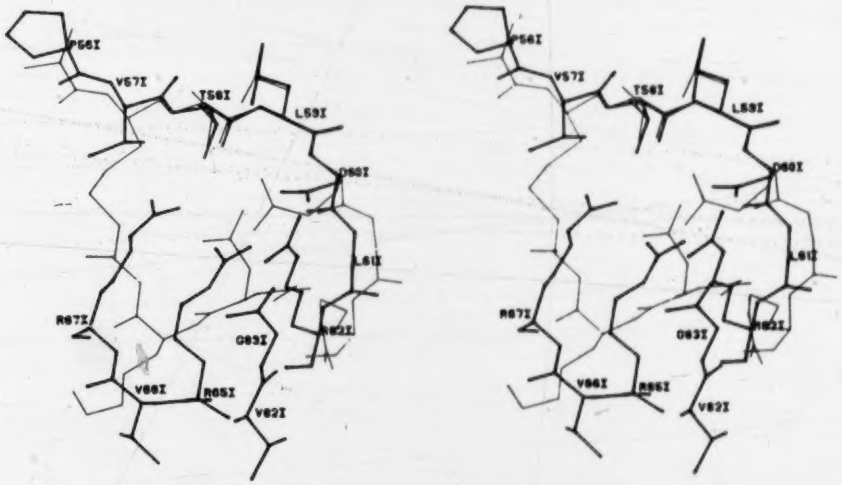

P56I
V57I
T58I
L59I
D60I
L61I
R67I
R62I
G63I
V66I
R65I
V62I
P56I
V57I
T58I
L59I
D60I
L61I
R67I
R62I
G63I
V66I
R65I
V62I

Figure 5.11. Superposition of Eglin-c and OMTKY3 Reactive Site Loops. The reactive site loops of the two inhibitors were overlapped as described in the text. Eglin-c is drawn with thick lines, OMTKY3 with thin lines. Residues of eglin-c are labelled.

Read *et al.* (1983) have proposed that barriers to relaxation of the reactive site region after hydrolysis would also contribute to the overall slow hydrolysis. They cite the disulphide bridges flanking the reactive site bond and involvement of the C-terminal end of the reactive site loop in β-structure as factors that would reduce the conformational freedom of the cleaved inhibitor. The covalent disulphide linkages are missing in CI-2 and eglin-c but may be replaced by the extensive hydrogen-bonding network stabilizing the reactive site loop. The residues before and after the loop in these two inhibitors are part of their central parallel β-sheet, again analogous to the interactions in OMTKY3.

Specific barriers to separation of the new termini in a cleaved inhibitor could be another factor contributing to slow hydrolysis. Figure 5.11 shows the overlapped reactive site regions of eglin-c and OMTKY3 (least-squares superposition of main-chain plus C^{β} atoms of P_2, P_1, and P_1', rms deviation of 0.24Å). The hydrogen-bonding and electrostatic interactions mentioned above would be unfavorably disrupted by separation of the termini. A means of stabilizing the charges on the new termini of a cleaved bond is common to the inhibitors. The side-chain of P_1' is bent down toward its own NH in the eglin-c complex, close to hydrogen-bonding position; this negatively charged side-chain (Asp in eglin-c, Glu in CI-2 and OMTKY3) is well situated to interact with a new positively charged

N-terminus. In eglin-c and CI-2, the negative charge on a new C-terminus would be readily compensated by the side-chains of Arg65I and Arg67I. The positive end of the helix dipole (Asn33I to Asn45I) in OMTKY3 lies on the tip of Arg65I from eglin-c when the two are overlapped, and the dipole points toward the reactive site bond.

A model of CI-2 after cleavage of the reactive site bond can be built into the active site of subtilisin Novo. Starting from the coordinates of the intact inhibitor, rotations about the N-C^{α} bond of P_1 and the C^{α}-C bond of P_1' are sufficient to move the new termini to non-bonded contact distance. The carbonyl carbon of P_1 can be placed about 4.8Å from the O^{γ} of Ser221, to simulate the acyl enzyme. The side-chain of Met59I can be adjusted to a more extended conformation within the broad S_1 specificity pocket. Movement of the NH of P_1' away from the carbonyl carbon of P_1 shifts the side-chain of P_1' away from its position in the hydrogen-bonding network, but rotations about χ_1 and χ_2 restore most of its interactions. No strong conclusions can be drawn from such a rough model, but it does indicate that the inhibitor could remain bound to the enzyme after cleavage without excessive disruption of favorable binding and stabilizing interactions.

Differences in the rate of dissociation of uncleaved and modified OMTKY3 from α-chymotrypsin but not from the subtilisins (Ardelt & Laskowski, 1985), probably are due to different interactions between the enzymes and inhibitor.

The most important differences must be in interactions on the C-terminal side of the scissile bond. Once the bond has been cleaved and an acyl enzyme complex formed, the P' segment of the reactive site loop must shift enough to allow access of a water molecule to the active site. The water molecule must attack the acyl enzyme complex before the cleaved inhibitor can be released from the enzyme. The freedom of the P' segment to adjust its position in the active site thus should have an influence on the rate of hydrolysis and release of the inhibitor from the enzyme.

A structure of OMTKY3 with subtilisin is not available but CI-2 in complex with subtilisin Novo is a reasonable model. The P_1' to P_3' residues are the same in both inhibitors and the conformations of their reactive site loops are similar (McPhalen *et al.*, 1985). The interactions of CI-2 with the active sites of subtilisin Novo and α-chymotrypsin overlapped with subtilisin Novo (as for subtilisin Carlsberg, see above) were examined on the graphics system. There are relatively few enzyme:inhibitor interactions on the P' side of the scissile bond (section 2.4.1) to interfere with movements of the P' segment of the inhibitor.

The most obvious difference between subtilisin and α-chymotrypsin in this region is the presence of an extra surface loop in α-chymotrypsin, residues Gln34 to Phe39. This loop lies close to the side-chain of the P_2' residue and has no counterpart in subtilisin. The loop may interfere with adjustment of the position of the P' segment with formation of the acyl enzyme and attack of a water molecule, or it may interfere with release of the cleaved inhibitor from the enzyme. This is only a speculation, though; it is difficult to make good comparisons of this nature among the two subtilisin complex structures and the two OMTKY3 complex structures due to the changes in orientation of the inhibitors relative to the enzymes. Until the causes of the different orientations are more firmly established, comparisons of contacts and interactions can be only tentative.

5.8 Future Directions

Comparison of enzyme:inhibitor interactions to find structural factors affecting kinetic and biochemical properties of protein inhibitors would be aided by the availability of more X-ray structures. Two structures that would be of particular interest with respect to some of the questions raised in this thesis are complexes of OMTKY3 with subtilisin, and eglin-c or CI-2 with α-chymotrypsin.

Kinetic parameters such as k_{cat} and K_{hyd} for CI-2 and eglin-c with a variety of serine proteinases are of interest; knowledge of these parameters would establish whether inhibitors from the PI-1 family really do function by the standard mechanism of Laskowski & Kato. Measurement of association rate constants for uncleaved and modified inhibitors other than OMTKY3 with serine proteinases might indicate whether the ratios observed by Ardelt & Laskowski (1985) are characteristic of the enzymes. Other inhibitors may have different relative ratios with the enzymes tested, and such information would indicate whether the ratios are dependent only on the enzyme structure or whether possible inhibitor contributions should also be examined.

Further model-building of enzyme:inhibitor complexes such as eglin-c with α-chymotrypsin (in the absence of crystals and an X-ray structure), perhaps combined with energy minimization of the models, might provide clues about additional structural features contributing to barriers to hydrolysis. Models could also be built of cleaved inhibitors in different enzyme active sites, to examine interactions affecting dissociation rate constants.

Calculations of the electrostatic field in various regions of the active site could quantify some of the electrostatic interactions proposed to contribute barriers to hydrolysis. In particular, the electrostatic field at the reactive site bonds of eglin-c and OMTKY3 due to the P_1' side-chain and the positive charges of the arginines (eglin-c) and the helix dipole (OMTKY3) would be interesting to compare. Another interesting comparison would be the electrostatic effect on the distribution of charge in His64 and Asp32 due to the different hydrogen bonds formed by the catalytic residues in α-chymotrypsin and the subtilisins.

Finally, no thesis on protein structure would be complete in this era without some suggestions for experiments using site-directed mutagenesis. In this area, we have been anticipated by some of the biotechnology companies working with subtilisin. The role of Asn155 as part of the oxyanion hole is being explored; the effects of mutations at this position on binding constants and rates of catalysis have been examined (D. Estell, personal communication). Substitutions at Met222 are also being characterized, particularly to find mutants less susceptible to oxidative inactivation in detergent solutions (Estell *et al.*, 1985). The specificity of subtilisin Novo has been altered with mutations at Gly166, at the bottom of the S_1 binding pocket (Wells *et al.*, 1985).

The role of Ca^{2+} in stabilizing subtilisi~~n~~s might be explored by mutations in the residues that act as ligands in Ca^{2+} ion site 1, such as Asp41 to Asn41 and/or Gln2 to Glu2. The importance of binding interactions on the P' side of the inhibitor scissile bond could be examined by substitutions such as Phe189 to Tyr189, Ile189, or less conservative changes to reduce the hydrophobic interactions. Mutation of Tyr217 to Glu217 might increase binding of inhibitors with Arg in the P_3 position (Table 2.11).

Mutations in CI-2 and eglin-c would also be of interest. The most obvious changes to make would be to residues forming the hydrogen-bonding network supporting the reactive site loop e.g. Glu60I or Asp60I to Gln60I or Asn60I, Thr58I to Ser58I or Val58I, and any of the 3 arginines to lysines or histidines. The effect of these relatively conservative changes on inhibitory properties could indicate the most important contributions to stability of the loop. The differences in packing between the hydrophobic cores of eglin-c and CI-2 could be examined by systematic mutation of one inhibitor core to the other. The importance of the position of the body of the inhibitor relative to the reactive site loop might be indicated by such mutations i.e. how much flexibility in the inhibitor can be tolerated.

Bibliography

Alden, R. A., Birktoft, J. J., Kraut, J., Robertus, J. D., & Wright, C. S. (1971) *Biochem. Biophys. Res. Commun. 45*, 337-344.

Ardelt, W. & Laskowski, M. Jr. (1985) *Biochemistry 24*, 5313-5320.

Barel, A. O. & Glazer, A. N. (1968) *J. Biol. Chem. 243*, 1344-1348.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. J. Jr., Brice, M. D., Rogers, J. K., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535-542.

Drenth, J., Hol, W. G., Jansonius, J. W., & Koekoek, R. (1972) *Eur. J. Biochem. 26*, 177-181.

Estell, D. A., Graycar, T. P., & Wells, J. A. (1985) *J. Biol. Chem. 260*, 6518-6521.

Finkenstadt, W. R. & Laskowski, M. Jr. (1967) *J. Biol. Chem. 242*, 771-773.

Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M. Jr., & James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868-4872.

Hirono, S., Akagawa, H., Mitsui, Y., & Iitaka, Y. (1984) *J. Mol. Biol. 178*, 389-413.

Hol, W. G. J. (1971) Ph. D. Thesis, University of Gröningen.

James, M. N. G., Brayer, G. D., Delbaere, L. T. J., Sielecki, A. R., & Gertler, A. (1980) *J. Mol. Biol. 139*, 423-438.

Kraut, J., Robertus, J. D., Birktoft, J. J., Alden, R. A., Wilcox, P. E., & Powers, J. C. (1972) *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 36, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 117-123.

McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242-7246.

Laskowski, M. Jr. & Kato, I. (1980) *Ann. Rev. Biochem. 49*, 593-626.

Quast, U., Engel, J., Steffen, E., Tschesche, H., &

Kupfer, S. (1978) *Eur. J. Biochem. 86*, 353-360.

Read, R. J. & James, M. N. G. (1986) in *Proteinase Inhibitors*, ed. A .J .Barrett, Amsterdam: Elsevier, in press.

Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420-4433.

Read, R., Fujinaga, M., Sielecki, A., Ardelt, W., Laskowski, M. Jr., & James, M. (1984) *Acta Cryst. A40*, C50-C51.

Robertus, J. D., Kraut, J., Alden, R. A., & Birktoft, J. J. (1972) *Biochemistry 11*, 4293-4303.

Schnebli, H. P., Seemüller, U., Fritz, H., Maschler, R., Liersch, M., Bodmer, J. L., Virca, G. D., Lucey, E. C., Stone, P. G., & Snider, G. L. (1985) in *Intracellular Protein Catabolism*, eds. E. A. Khairallah, J. S. Bond, & J. W. C. Bird, New York: Alan Liss, Inc., 287-290.

Svendsen, I., Boisen, S. & Hejgaard, J. (1982) *Carlsberg Res. Commun. 47*, 45-53.

Wells, J. A., Powers, D. B., Miller, J., Graycar, T., & Estell, D. A. (1985) *Journal of Cellular Biochemistry 9B*, 98.

Wright, C. S. (1972) *J. Mol. Biol. 67*, 151-163.

Wright, C. S., Alden, R. A., & Kraut, J. (1969) *Nature (London) 221*, 235-242.